

Physics 221A
Fall 2020
Notes 2
The Postulates of Quantum Mechanics†

1. Introduction

In these notes we present the postulates of quantum mechanics, which allow one to connect experimental results with the mathematical formalism described in Notes 1. Actually, we are not ready to state the postulates in their complete and final form, since that requires the use of the density operator, which we discuss in Notes 3. Therefore our first version of the postulates will involve some undefined terminology, and will be incomplete. Nevertheless, even in their incomplete form, the postulates explain a good deal about the mathematical formalism of quantum mechanics and its relation to physical reality.

These postulates are not supposed to be obvious. In fact, even after the correct equations describing quantum mechanics had been discovered (by Heisenberg and Schrödinger, in 1925–1926), it was still not obvious what the correct interpretation should be. The interpretation presented in these notes and the next set is the orthodox one, although there are controversial aspects about it and there are alternatives that are active competitors even today. For example, recent work in cosmology has led to renewed interest in the “many worlds” interpretation, in which the “wave function of the universe” plays a role. For now it is best to state the postulates of the orthodox interpretation and see how they work out on some specific physical examples.

2. Postulates of Quantum Mechanics (Incomplete Form)

In their incomplete form, the postulates of quantum mechanics are the following:

1. Every physical system is associated with a Hilbert space \mathcal{E} . We call the vectors of this space *kets*.
2. Every pure state of a physical system is associated with a definite ray in \mathcal{E} . We postpone for a while the definition of a pure state, but a ray was defined in Eq. (1.9). In practice we often represent the state of a system by some nonzero ket lying in the ray in question. In this course we will often refer to a nonzero ket $|\psi\rangle$ as the “state” of the system, but really it is the ray that corresponds to the state (the normalization and phase of the ket are of no physical significance).

† Links to the other sets of notes can be found at:

<http://bohr.physics.berkeley.edu/classes/221/1920/221.html>.

3. Every measurement that can be carried out on the system corresponds to a complete Hermitian operator A (an observable).
4. The possible results of the measurement are the eigenvalues of A , either the discrete eigenvalues a_1, a_2, \dots or the continuous ones $a(\nu)$.
5. In the discrete case, the probability of measuring $A = a_n$ is

$$\text{Prob}(A = a_n) = \frac{\langle \psi | P_n | \psi \rangle}{\langle \psi | \psi \rangle}, \quad (1)$$

where P_n is the projection operator onto the eigenspace \mathcal{E}_n corresponding to eigenvalue a_n , as indicated by Eq. (1.124), and where $|\psi\rangle$ is any nonzero ket in the ray representing the state of the system. In the continuous case, the probability of measuring A to lie in some interval $I = [a_0, a_1]$ of the continuous spectrum is

$$\text{Prob}(a_0 \leq A \leq a_1) = \frac{\langle \psi | P_I | \psi \rangle}{\langle \psi | \psi \rangle}, \quad (2)$$

where P_I is the projection operator corresponding to interval I , as in Eq. (1.128). Notice that the probabilities in Eqs. (1) and (2) depend only on the ray, not on which ket is chosen to represent the ray.

6. After a measurement with discrete outcome $A = a_n$, the system is represented by the ket $P_n|\psi\rangle$, where $|\psi\rangle$ represents the system before the measurement. Note that in this case, the system is in an eigenstate of A with eigenvalue a_n after the measurement. In the continuous case, with outcome $a_0 \leq A \leq a_1$, the system is represented by $P_I|\psi\rangle$ after the measurement, where again P_I is as in Eq. (1.128).

As postulate 5 shows, the predictions made by quantum mechanics are of a statistical nature. The probabilities referred to in that postulate are determined experimentally by repeating the same measurement on a large number of identically prepared systems. The fact that identically prepared systems can yield different results on successive measurements is a point that we will discuss after we give some examples.

We emphasize that quantum mechanics makes statistical predictions about measurements on an *ensemble*, that is, a collection of (conceptually) an infinite number of identically prepared systems. Although we often use language like “the wave function of the electron,” in reality the state vector (or wave function) does not describe an individual system, but only the statistics of an ensemble. And postulate 6, also called the *collapse* postulate, does not refer to the “collapse” of any individual system, but rather the creation of a new ensemble, described by a new state vector, that results from the act of making a measurement on the old ensemble.

There is a seventh postulate, not listed here, which states that the state vector evolves in time by means of a time-dependent, unitary transformation. This subject will be taken up in Notes 5. Postulate 7 is actually in conflict with postulate 6, assuming that the measuring apparatus is governed by the same rules of quantum mechanics as the system being measured. This is because

the projection required by postulate 6 cannot be achieved by unitary transformations, nor can the measurement apparatus end up in a definite state (indicating the result of the measurement) by such means. Much of the ongoing discussion and debate regarding the quantum theory of measurement involves such issues as these in an attempt to explain how in practice we do end up with definite measurements. An aspect that cannot be neglected is that a real measurement apparatus is always macroscopic object that cannot be isolated from its environment. Thus any consideration of the quantum state of the apparatus must take into account the interactions with the environment in some statistical way.

We could pursue these questions but for now the more important task is to see how these postulates actually work in practice. For example, we will have to answer questions such as the following: How do we know what Hilbert space is associated with a given physical system? What operator corresponds to definite measurements? And which ray corresponds to a definite (pure) state of system?

3. Postulates Applied to the Stern-Gerlach Experiment

To answer such questions, we will analyze the Stern-Gerlach experiment, pretending that we know nothing except the experimental results and the postulates listed above. In particular, we will pretend that we know nothing about wave functions, spin, Pauli matrices, the Schrödinger equation, etc. The following presentation is based closely on that given by J. J. Sakurai, *Modern Quantum Mechanics*.

Suppose we are working with a beam of silver atoms, as in the original Stern-Gerlach experiment. From a modern perspective, we know that silver atoms possess a magnetic moment $\boldsymbol{\mu}$ because of their single unpaired valence electron, so that the measured value of any component of magnetic moment is $\pm\mu_0$, where $\mu_0 = e\hbar/2mc$ is a Bohr magneton (this is in Gaussian units; see Appendix A). The silver atom has the same electronic spin and magnetic moment as a free electron, but it is electrically neutral. Charged particles are not suitable for a Stern-Gerlach experiment, because in practice the ordinary Lorentz force $(q/c)\mathbf{v}\times\mathbf{B}$ is much larger than the force due to the magnetic moment, $\mathbf{F} = \nabla(\boldsymbol{\mu} \cdot \mathbf{B})$. The original experiment of Stern and Gerlach in 1921 provided a measurement of μ_0 , with a result in good agreement with Bohr's value. We also know that the magnetic moment operator is proportional to the spin operator, $\boldsymbol{\mu} = (e/mc)\mathbf{S}$, where any component of spin takes on the values $\pm\hbar/2$. In the following discussion, however, we will play dumb and ignore all of this, and instead we will work solely with the experimental results. For the same reason, we will speak in terms of measurements of magnetic moment, not spin.

Consider the tandem Stern-Gerlach apparatus illustrated in Fig. 1, in which we first measure the x -component of the magnetic moment $\boldsymbol{\mu}$, and then pass the beam with $\mu_x = +\mu_0$ to a second magnet which measures μ_z . The experimental result is that the two beams emerging from the second magnet, with $\mu_z = \pm\mu_0$, each carry 50% of the atoms that entered that magnet.

Since a measurement of μ_x gives rise to two possible outcomes, we must assume according to

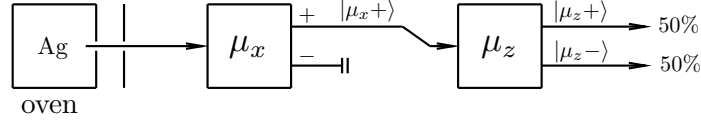


Fig. 1. A beam of silver atoms is subjected to a measurement of μ_x , after which the atoms with $\mu_x = +\mu_0$ are passed to a second magnet which measures μ_z .

postulates 3 and 4 that the operator μ_x has two possible eigenvalues, $\pm\mu_0$. The same is true for μ_y and μ_z , or indeed, for any component of spin (in any direction). Therefore the Hilbert space \mathcal{E} must be at least 2-dimensional, since the eigenspaces corresponding to $\mu_x = \pm\mu_0$ are orthogonal. Indeed, measurements of the components of the magnetic moment alone give us no indication that the Hilbert space has more than two dimensions. The easiest way to proceed with the following argument is simply to assume that the eigenvalues $\pm\mu_0$ are nondegenerate, and then to come back later to the question of degeneracies. Under this assumption, the eigenspaces are 1-dimensional, the ket space \mathcal{E} has exactly two dimensions, and it is spanned by the eigenkets of μ_x with eigenvalues $\pm\mu_0$. It is also spanned by eigenkets of μ_y or μ_z with eigenvalues $\pm\mu_0$, and therefore the pair of eigenkets of any one of these operators must be expressible as linear combinations of the eigenkets of any other operator.

Let us denote some normalized eigenvector of μ_x with eigenvalue $+\mu_0$ by $|\mu_x+\rangle$; we say “some” because any other vector differing from this one by an overall phase would work just as well at this stage of the argument. Similarly, let us choose normalized eigenvectors $|\mu_x-\rangle$, $|\mu_y\pm\rangle$, and $|\mu_z\pm\rangle = |\pm\rangle$ (we will henceforth omit the μ_z specification in the eigenkets of this operator). Because eigenkets corresponding to distinct eigenvalues are orthogonal, we have

$$\langle\mu_x+|\mu_x-\rangle = \langle\mu_y+|\mu_y-\rangle = \langle+|- \rangle = 0. \quad (3)$$

In accordance with postulate 2, the state of the atomic beam at various stages in the apparatus is described by some state vector (as we will discuss later, the spin system is in a pure state anywhere after the measurement of μ_x in the first magnet). Also, according to postulate 6, after the measurement of μ_x with result $+\mu_0$, the state of the system is specified by $|\mu_x+\rangle$, as indicated by the ket attached to the upper beam in Fig. 1. This is because by our assumption the eigenspace of μ_x with eigenvalue μ_0 is one-dimensional, that is, it is a ray, so the projection operator onto this subspace produces a vector in that ray. The state of the system is determined by the ray, which is specified by any nonzero vector that lies in it, such as $|\mu_x+\rangle$. Similarly, the two beams emerging from the second magnet, which measures μ_z , are represented by the state vectors $|\mu_z\pm\rangle = |\pm\rangle$. We represent the state $|\mu_x+\rangle$ entering the second magnet as a linear combination of the eigenkets $|\pm\rangle$,

$$|\mu_x+\rangle = c_+|+\rangle + c_-|-\rangle, \quad (4)$$

where c_{\pm} are the expansion coefficients. Since $|+\rangle$ and $|-\rangle$ are orthogonal, this implies

$$c_{\pm} = \langle\pm|\mu_x+\rangle. \quad (5)$$

Then, according to postulate 5, we have

$$\text{Prob}(\mu_z = +\mu_0) = \langle \mu_x + | P_+ | \mu_x + \rangle, \quad (6)$$

where

$$P_+ = |+\rangle\langle +| \quad (7)$$

is the projection operator onto the $+\mu_0$ eigenspace of μ_z . The denominator in Eq. (2) is unity because $|\mu_x +\rangle$ is normalized. Therefore Eq. (6) becomes

$$\text{Prob}(\mu_z = +\mu_0) = \langle \mu_x + | + \rangle \langle + | \mu_x + \rangle = |\langle + | \mu_x + \rangle|^2 = |c_+|^2 = \frac{1}{2}, \quad (8)$$

where we use Eq. (4) and the experimentally measured probability. This implies

$$c_+ = \frac{1}{\sqrt{2}} e^{i\alpha_1}, \quad (9)$$

where $e^{i\alpha_1}$ is an unknown phase factor. But since $\text{Prob}(\mu_z = -\mu_0)$ is also $1/2$, a similar argument gives

$$c_- = \frac{1}{\sqrt{2}} e^{i\beta_1}, \quad (10)$$

where $e^{i\beta_1}$ is another unknown phase. Thus, we have

$$|\mu_x +\rangle = \frac{1}{\sqrt{2}} (e^{i\alpha_1} |+\rangle + e^{i\beta_1} |-\rangle). \quad (11)$$

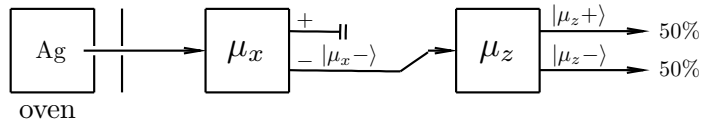


Fig. 2. Same as Fig. 1, except the beam with $\mu_x = -\mu_0$ is fed into the second magnet.

In general, the *wave function* of a system is the expansion coefficients of the state vector with respect to some specified basis. Since the coefficients c_+ and c_- are the expansion coefficients of $|\mu_x +\rangle$ with respect to the eigenbasis of μ_z , they constitute the wave function in the μ_z -representation. We see that the analysis so far, which makes use of the measured probabilities in Fig. 1, has determined the wave function (c_+, c_-) only to within the two phases shown in Eq. (11). We will now pin down these phases by using a combination of further experimental data and phase conventions.

Equation (11) can be simplified by changing the phase of the ket $|\mu_x +\rangle$ so as to absorb the phase $e^{i\alpha_1}$ on the right hand side. Let $|\mu_x +\rangle = |\mu_x +\rangle' e^{i\alpha_1}$ and multiply Eq. (11) by $e^{-i\alpha_1}$, setting $\beta'_1 = \beta_1 - \alpha_1$. Then we have

$$|\mu_x +\rangle = \frac{1}{\sqrt{2}} (|+\rangle + e^{i\beta_1} |-\rangle), \quad (12)$$

after dropping the primes. Now we have established a phase convention for the ket $|\mu_x+\rangle$ and we cannot change its phase again. By a similar analysis, if we feed the beam emerging from the first magnet with $\mu_x = -\mu_0$ into the second magnet, as illustrated in Fig. 2, and analyze the probabilities exactly as we have done here, we obtain

$$|\mu_x-\rangle = \frac{1}{\sqrt{2}}(|+\rangle + e^{i\gamma_1}|-\rangle), \quad (13)$$

where $e^{i\gamma_1}$ is another phase. Now we have fixed the phases of both kets $|\mu_x\pm\rangle$.

But the phases $e^{i\beta_1}$ and $e^{i\gamma_1}$ are related, for by orthogonality we have

$$\langle\mu_x+|\mu_x-\rangle = \frac{1}{2}[1 + e^{i(-\beta_1+\gamma_1)}] = 0, \quad (14)$$

or $e^{i\gamma_1} = -e^{i\beta_1}$. Thus, Eqs. (12) and (13) can be written

$$|\mu_x\pm\rangle = \frac{1}{\sqrt{2}}(|+\rangle \pm e^{i\beta_1}|-\rangle). \quad (15)$$

Similarly, suppose we measure μ_y instead of μ_x in the first magnet in Figs. 1 and 2. Then the two beams that emerge from the second magnet still carry 50% each of the silver atoms, and an analysis just like that presented above gives

$$|\mu_y\pm\rangle = \frac{1}{\sqrt{2}}(|+\rangle \pm e^{i\beta_2}|-\rangle), \quad (16)$$

where $e^{i\beta_2}$ is another phase. Now the phases of kets $|\mu_y\pm\rangle$ have been fixed.

We can now write down the operators representing the components of $\boldsymbol{\mu}$ in the $|\pm\rangle$ basis. For example, according to Eq. (1.127) we have

$$\mu_x = \mu_0 \left(|\mu_x+\rangle\langle\mu_x+| - |\mu_x-\rangle\langle\mu_x-| \right), \quad (17)$$

which can be transformed to the $|\pm\rangle$ basis by substituting Eq. (15). This gives

$$\mu_x = \mu_0 \left(e^{-i\beta_1}|+\rangle\langle-| + e^{i\beta_1}|-\rangle\langle+| \right). \quad (18)$$

Similarly, we find

$$\mu_y = \mu_0 \left(e^{-i\beta_2}|+\rangle\langle-| + e^{i\beta_2}|-\rangle\langle+| \right), \quad (19)$$

and of course we have

$$\mu_z = \mu_0 \left(|+\rangle\langle+| - |-\rangle\langle-| \right). \quad (20)$$

Next, a relation can be found between the phases $e^{i\beta_1}$ and $e^{i\beta_2}$ by imagining another Stern-Gerlach experiment, like Fig. 1 except that we measure μ_y in the second magnet instead of μ_z . Again, we can take it as an experimental result that the two values $\mu_y = \pm\mu_0$ are measured with equal probability. Therefore we have

$$\frac{1}{2} = |\langle\mu_y\pm|\mu_x+\rangle|^2 = \frac{1}{2}[1 \pm \cos(\beta_2 - \beta_1)], \quad (21)$$

where we use Eqs. (15) and (16). But this implies $\beta_2 = \beta_1 \pm \pi/2$, or

$$e^{i\beta_2} = \pm i e^{i\beta_1}. \quad (22)$$

Thus, the two unknown phases in Eqs. (18) and (19) reduce to one unknown phase and an unknown sign.

The remaining unknown phase, say, $e^{i\beta_1}$, can be pinned down by choosing a phase convention for the ket $|-\rangle$; for example, we can make the matrix elements of either μ_x or μ_y in the $|\pm\rangle$ basis purely real, but not both, because by Eq. (22), if the matrix elements of one of these operators is real, those of the other operator must be purely imaginary. We see that it is impossible to satisfy the postulates of quantum mechanics and the experimental results with only real numbers; complex numbers are necessary. Here we choose to absorb the phase $e^{i\beta_1}$ into the definition of the ket $|-\rangle$, so that Eqs. (18), (19) and (20) become

$$\begin{aligned} \mu_x &= \mu_0 (|+\rangle\langle -| + |-\rangle\langle +|), \\ \mu_y &= \pm \mu_0 (-i|+\rangle\langle -| + i|-\rangle\langle +|), \\ \mu_z &= \mu_0 (|+\rangle\langle +| - |-\rangle\langle -|). \end{aligned} \quad (23)$$

We notice that we cannot change the phase of $|+\rangle$ without messing up equations such as (15) or (16). Thus, all the freedom in phase conventions has been used up. To state results equivalent to Eqs. (23), we can write out the matrices representing the three operators in the $|\pm\rangle$ basis,

$$\mu_x = \mu_0 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mu_y = \pm \mu_0 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \mu_z = \mu_0 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (24)$$

where the basis vectors are ordered $|+\rangle, |-\rangle$. We frequently write equations like (24), but we do not really mean that the operator on the left is equal to the matrix on the right, instead we mean that its matrix elements in some given or understood basis are given by the matrices. Here the basis is the eigenbasis of μ_z . We see in Eq. (24) the appearance of the Pauli matrices.

We see that to within a final unknown sign, the three Pauli matrices are determined solely from the postulates of quantum mechanics, the experimental results, and some phase conventions. But there is the question of the final, unknown sign. Sakurai has some discussion of this point, but he never definitively settles the issue. We will look at this question more carefully in Prob. 2.

4. Compatible or Commuting Observables

So far we have simply assumed that the eigenspaces of μ_x (and μ_y, μ_z) are nondegenerate. How would we know if this were not true? If we were given a Hermitian operator or matrix as a purely mathematical problem, then the answer could be obtained purely by mathematics: we first compute the eigenvalues, and then the order of the degeneracy of an eigenvalue is the number of linearly independent eigenvectors of that eigenvalue. But here we are building up the Hilbert space out of

the results of physical measurements, and there must be physical meaning to any degeneracies that might exist. As we will now show, the answer to this question involves the notion of compatible or commuting observables.

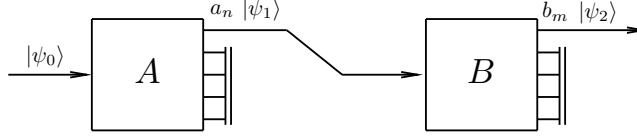


Fig. 3. A quantum system is subjected to a measurement of two observables, first A , and then B .

Consider an idealized measurement such as illustrated in Fig. 3. A system that is known to be in a pure state $|\psi_0\rangle$ is first subjected to a measurement of observable A . Out of the several possible outcomes, all are thrown away except a_n . The system after the measurement of $A = a_n$ is described by $|\psi_1\rangle$; this system is passed to a device that measures observable B , and all outcomes except $B = b_m$ are thrown away. The state of the system after the second measurement is described by $|\psi_2\rangle$.

According to the postulates, the probability of measuring $A = a_n$ in the first apparatus is

$$\text{Prob}(a_n) = \frac{\langle \psi_0 | P_{A_n} | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle}, \quad (25)$$

where P_{A_n} is the projection operator onto the eigenspace of A corresponding to eigenvalue a_n . Also, the ket describing the state that emerges from the first filter is

$$|\psi_1\rangle = P_{A_n} |\psi_0\rangle. \quad (26)$$

Next let us compute the probability of measuring first $A = a_n$ and then $B = b_m$. This is the (conditional) probability of measuring $B = b_m$ given that we had $A = a_n$, multiplied times the probability of measuring $A = a_n$ in the first place. We denote the probability of this compound measurement by $\text{Prob}(a_n; b_m)$; it is given by

$$\text{Prob}(a_n; b_m) = \frac{\langle \psi_1 | P_{B_m} | \psi_1 \rangle}{\langle \psi_1 | \psi_1 \rangle} \frac{\langle \psi_0 | P_{A_n} | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle}, \quad (27)$$

where P_{B_m} is the projection operator onto the eigenspace of B with eigenvalue b_m . But we also have

$$\langle \psi_1 | \psi_1 \rangle = \langle \psi_0 | P_{A_n}^\dagger P_{A_n} | \psi_0 \rangle = \langle \psi_0 | P_{A_n} | \psi_0 \rangle, \quad (28)$$

where we have used Eq. (26) and the fact that P_{A_n} is Hermitian and idempotent. Thus, the denominator of the first factor in Eq. (27) cancels the numerator of the second factor, and we have

$$\text{Prob}(a_n; b_m) = \frac{\langle \psi_0 | P_{A_n} P_{B_m} P_{A_n} | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle}. \quad (29)$$

Similarly, if we reverse the order of the measurements, measuring first $B = b_m$ and next $A = a_n$, we obtain the probability

$$\text{Prob}(b_m; a_n) = \frac{\langle \psi_0 | P_{Bm} P_{An} P_{Bm} | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle}. \quad (30)$$

These two probabilities are not equal, in general.

Now it can be shown that if $[A, B] = 0$, then $[P_{An}, P_{Bm}] = 0$, so in this case we have

$$P_{An} P_{Bm} P_{An} = P_{An}^2 P_{Bm} = P_{An} P_{Bm} = P_{Bm} P_{An} P_{Bm}, \quad (31)$$

so that both probabilities are equal,

$$\text{Prob}(a_n; b_m) = \text{Prob}(b_m; a_n) = \frac{\langle \psi_0 | P_{An} P_{Bm} | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle}. \quad (32)$$

Conversely, it can be shown that if the probabilities $P(a_n; b_m)$ and $P(b_m; a_n)$ are equal for all initial states $|\psi_0\rangle$ and all n and m , then $[A, B] = 0$.

Altogether, we see that probability of a compound measurement is independent of the order of the measurements, for all initial states and all outcomes of the two measurements, if and only if the observables commute. This is the physical meaning of commuting observables. Commuting observables are also said to be *compatible*. A lesson of this analysis is that it is in principle an experimental matter to determine if two observables, corresponding to definite measurement processes, commute; we simply carry out those measurements in different orders on a large number of systems, and compare the resulting probabilities.

We have described an example of what we might call *quantum statistics*, in which probabilities depend on the order in which measurements are carried out. A similar problem in *classical statistics*, by which we mean ordinary statistics, is the following. Suppose we have an urn containing balls with spots of various colors painted on them. If we first select the balls that have a red spot, and then feed these to a device that selects balls that have a green spot, then what comes out is the set of balls that have both a red and a green spot. The number of these balls, divided by the total number of balls in the urn, is the probability of selecting a ball with both a red and a green spot from the urn in the first place. Moreover, the same number is obtained if we select balls with green spots first and then those with red spots. The order does not matter. In this sense, classical statistics is commutative, while quantum statistics is noncommutative. We will have more to say later about the dependence in quantum mechanics of experimental results on the order in which measurements are carried out.

5. Resolving Degeneracies; Complete Sets of Commuting Observables

Let us now return to the question of degeneracies. Referring to Fig. 3, how do we know if the eigenvalues of the operator A are degenerate? (You may wish to review the proof of the theorem that commuting observables possess a simultaneous eigenbasis, given in Sec. 1.23, to better understand this section.) If an eigenvalue a_n is degenerate, then the eigenspace \mathcal{E}_n corresponding

to this eigenvalue is multidimensional, so we are asking about the dimensionality of the subspaces \mathcal{E}_n . The answer is obtained by searching for other observables B that commute with A , to see if one of them will ‘resolve’ the degeneracy of a_n , that is, produce more than one outcome when a measurement of B is made subsequent to the measurement $A = a_n$. If such an observable can be found, then the states emerging from the B -apparatus lie in the simultaneous eigenspaces of the operators A and B , which are subspaces of the eigenspace \mathcal{E}_n of A . A vector lying in one of these subspaces is obtained by applying the projectors P_{An} and P_{Bm} (in either order, since they commute) to an arbitrary ket, such as $|\psi_0\rangle$ in the figure. In this case, the order of the degeneracy of a_n is at least equal to the number of outcomes of the subsequent B -measurement, since each of the simultaneous eigenspaces is at least one-dimensional.

However, these simultaneous eigenspaces of A and B may themselves be degenerate. To find out if they are, we can search for another operator C that commutes with both A and B , that will resolve the simultaneous eigenspaces of A and B into smaller subspaces. The process continues until no more resolutions are possible; then the set of observables (A, B, C, \dots) constitutes a *complete set of commuting observables*, or CSCO for short. At this point we can declare that the simultaneous eigenspaces of the CSCO are nondegenerate, and the dimensionalities of all simultaneous eigenspaces of all operators in the CSCO are determined.

Once we have found a CSCO, then any new observable that commutes with the CSCO must be a function of the members of the CSCO. This is because the new observable does not produce any further resolutions of the simultaneous eigenspaces of the CSCO, so its eigenvalue on any one of these (one-dimensional) subspaces is just a number that can be regarded as a function of the eigenvalues of the members of the CSCO. This is what we mean by a function of an operator (see Sec. 1.25).

How do we know if we have gone far enough in the search for operators when trying to construct a CSCO? In short, we don’t. If we cannot find a new operator (that is, a new measurement process) that commutes with the operators already in the list and that creates a further resolution of the eigenspaces, it may be due to the lack of such an operator, or the lack of our imagination in searching for one. In practice, this means that there are degrees of freedom that we are ignoring, either because we don’t know about them or because they are not important for the physics we are interested in. For example, in the usual discussion of the hydrogen atom we ignore the degrees of freedom associated with the quarks in the nucleus.

In other words, we often choose not to go all the way to a complete set of commuting observables because we are only interested in some subset of the degrees of freedom of a system. For example, in our discussions of the Stern-Gerlach apparatus, we have been ignoring the spatial degrees of freedom of the silver atom, as well as the internal degrees of freedom representing the motion of the electrons around the nucleus, the motion of the nucleons in the nucleus, the quarks in the nucleons, etc. If we were to include the spatial degrees of freedom of the silver atoms as well as the spin, our Hilbert space would be infinite-dimensional (actually, $2 \times \infty$).

For another example, we frequently treat electrons as spinless particles (effectively ignoring the spin degrees of freedom), so that the wave function is a scalar $\psi(\mathbf{r})$ instead of a 2-component spinor. This is physically justified in the common case that the electron is moving in an electrostatic field, since spins do not interact with electric fields. It may also be useful as an approximation when spin-dependent effects are small, as they usually are in nonrelativistic quantum mechanics. Later, if we incorporate spin in a more realistic model, we will find that the Hilbert space becomes enlarged and new quantum numbers appear.

6. Noncommuting Measurements

As discussed above, it is a feature of quantum mechanics that the outcome of measurements depends in general on the order in which they are taken. This is sometimes described by saying that when we measure one observable, we introduce an unavoidable and uncontrollable disturbance in the values of all other observables that do not commute with the one we are measuring. For example, when measuring μ_z , we disturb the values of μ_x and μ_y , or when measuring the position of a particle we disturb its momentum. In comparison to the filtering of classical ensembles, for example, selecting balls from an urn with different colored spots on them, it is as if when we select a ball and discover it has a red spot, this fact alone causes all the other spots on the ball to be replaced by new ones with randomly chosen colors.

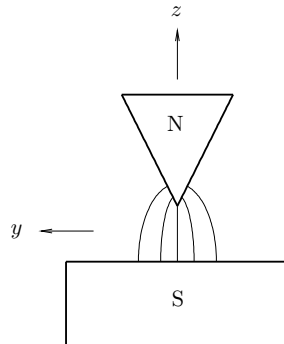


Fig. 4. The Stern-Gerlach apparatus seen from one end.

Such disturbances in noncommuting variables also occur in classical mechanics, it is just that in the classical framework we can in principle reduce the disturbance to an arbitrarily small value. To understand this point, let us consider the measurement of μ_z in a Stern-Gerlach apparatus in more detail, adopting a classical point of view. Measuring the force on an object in an inhomogeneous magnetic field in order to measure its magnetic moment is something we can do with any magnetized object, including macroscopic ones for which classical mechanics is valid.

An end-view of the apparatus is sketched in Fig. 4. The beam is directed along the x -axis and runs just below the knife edge of the upper pole tip, where the magnetic field has the largest gradient.

The force on the atoms is small, so the magnet extends as far as practicable in the x -direction, to maximize the momentum transfer in the z -direction. Obviously this momentum transfer must be greater than the statistical spread in the beam in p_z in order to see the effect at all. Assuming the beam is of small spatial extent, we can assume the magnetic field is purely in the z -direction at the location of the beam (directly below the pole tip), $\mathbf{B} = B_z \hat{\mathbf{z}}$, so $B_x = B_y = 0$ at the location of the beam. Actually, ignoring fringe fields, $B_x = 0$ everywhere due to symmetry. Both the z - and y -components of \mathbf{B} have nonzero gradients, however, at the location of the beam (these gradients are restricted by $\nabla \cdot \mathbf{B} = 0$). The force is $\mathbf{F} = \nabla(\boldsymbol{\mu} \cdot \mathbf{B})$. The magnetic moment $\boldsymbol{\mu}$ characterizes the magnetic particle and is a function of time, but it does not depend on space so the force can be written

$$\mathbf{F} = \mu_y \nabla B_y + \mu_z \nabla B_z. \quad (33)$$

The magnetic moment itself is not constant in the magnetic field, but precesses according to

$$\dot{\boldsymbol{\mu}} = \boldsymbol{\omega} \times \boldsymbol{\mu}, \quad (34)$$

where $\boldsymbol{\omega}$ is in the direction of \mathbf{B} . For the silver atom, $\boldsymbol{\omega}$ has the magnitude

$$\omega = \frac{eB}{mc}. \quad (35)$$

Equation (34) applies to a silver atom, but would not be true for a macroscopic magnetized object unless the magnetic moment were proportional to the angular momentum which is assumed to be large. Otherwise a more complicated equation of motion for $\boldsymbol{\mu}$ must be used. Of course Stern and Gerlach did not know this about silver atoms, but for simplicity we will content ourselves with the simple equation of motion (34), which is correct for silver atoms.

The consequence of Eq. (34) is that $\boldsymbol{\mu}$ precesses about the direction \mathbf{B} with frequency ω , sweeping out a cone, so that its component along \mathbf{B} (in this case, μ_z) is constant, while the component of $\boldsymbol{\mu}$ perpendicular to \mathbf{B} rotates in a circle with frequency ω . In fact, in a real Stern-Gerlach experiment, the atom stays in the magnetic field for a time T that is much greater than the precession period $\tau = 2\pi/\omega$. This is because, as mentioned, it is desirable to keep the particle in the inhomogeneous magnetic field as long as possible to maximize the momentum transfer.

Due to the precession, the component μ_y in Eq. (33) averages to zero on time scales long compared to the precession time, and the average force felt by the particle is

$$\mathbf{F} = \mu_z \nabla B_z = \mu_z \frac{\partial B_z}{\partial z} \hat{\mathbf{z}}, \quad (36)$$

where we use the fact that near the pole tip the gradient of B_z is predominantly in the z -direction. Thus, the momentum transfer in the z -direction is

$$\Delta p_z = T \mu_z \frac{\partial B_z}{\partial z}, \quad (37)$$

assuming $\partial B_z / \partial z$ does not change much as the beam gradually bends in response to the force (that is, assuming the spatial displacement in the z -direction is small enough, something that may or may not be true in practice).

An important point of this analysis is that while we do measure the (constant) value of μ_z while the atom is in the inhomogeneous magnetic field, the other two components μ_x and μ_y undergo precession and so are not constant. This happens even in a classical description, and is an example of the fact that measuring an observable introduces a disturbance in other observables that do not commute with it. In classical mechanics, however, it is possible in principle to account for the evolution of μ_x and μ_y in the apparatus. For any given particle with definite initial position and momentum, we can in principle calculate its trajectory through the magnet, integrate Eq. (34) along the trajectory, and compensate for the evolution of μ_x and μ_y . Of course, the actual beam has some spatial extent, so different particles will follow slightly different trajectories and the spins will precess by different total angles, producing a spread in the distribution of precession angles when the beam emerges. If this spread is much greater than 2π , then we will have the appearance even in classical mechanics of a random disturbance in μ_x and μ_y upon measuring μ_z .

In classical mechanics we can minimize this spread by making the beam sufficiently narrow in its spatial extent. In quantum mechanics, however, this is not possible because of the uncertainty principle applied to the position and momentum of the particles in the beam. An analysis of these restrictions shows that it is impossible to reduce the spatial extent of the beam enough to make the spread in precession phase angles much less than 2π . Thus, in quantum mechanics, there is indeed an unavoidable and uncontrollable disturbance in μ_x and μ_y , upon measuring μ_z . This question is examined further in Prob. 3.

Perhaps you have seen a discussion of the “Heisenberg microscope,” which is often used to explain the position-momentum uncertainty relations. This is a gedankenexperiment that attempts to measure the position of a particle with some accuracy by shining light on it, and forming the image with a lens. The position of the image is limited by diffraction of the light, which can be reduced by using light of shorter wavelength. This, however, increases the momentum transfer when the photon scatters off the particle we are trying to observe, introducing an unavoidable and uncontrollable disturbance in the momentum of the particle. The details of the Heisenberg microscope are different but the principle is the same as in the spin example we are discussing here.

In discussions of this sort we are really using a mixture of classical and quantum reasoning to understand certain phenomena. There is no harm in this, and classical reasoning must never be underestimated as a means for understanding quantum phenomena. Nevertheless, the kind of language we have been using ascribes a reality to certain quantities that are not observable, for example, the value of μ_x of the atom inside the Stern-Gerlach apparatus. It is a safer point of view, and more in accordance with the orthodox interpretation of quantum mechanics, that unobservable quantities should not be considered a part of physical reality. More on this point later.

7. The Ensemble Interpretation

You may be familiar with the “Schrödinger cat” paradox. Like Einstein, Schrödinger was a physicist who did not like the statistical interpretation of quantum mechanics. To argue against it,

he pointed out that linear combinations of microscopic quantum states could apparently be translated into linear combinations of macroscopic states, something we do not observe. His example was a (poor) cat confined to a box with in which there was a radioactive nucleus. As time goes on, the state of the nucleus becomes a linear combination of an excited nucleus and a nucleus in the ground state with an emitted particle. A detector sits nearby, and when it detects the emitted particle, it releases a capsule of cyanide which kills the cat. Now, Schrödinger asks, do we have a linear combination of a live cat and a dead cat?

We emphasize that quantum mechanics makes statistical predictions about measurements on an *ensemble*, that is, a collection of (conceptually) an infinite number of identically prepared systems. Although we often use language like “the wave function of the electron,” in reality the state vector (or wave function) does not describe an individual system, but only the statistics of an ensemble. And postulate 6, the collapse postulate, does not refer to any collapse of any individual system, but rather the creation of a new ensemble, described by a new state vector, that results from the act of making a measurement on the old ensemble. In this way we avoid apparent paradoxes of the “Schrödinger cat” type. That is, at any given time we have an ensemble of cats in boxes; if we pick one out and look at it (observe it), we find either a live cat or a dead one. Quantum mechanics correctly predicts the statistical distribution of these results.

To analyze the Schrödinger cat situation more carefully, one would have to take into account that the cat in the box is a macroscopic system, whose quantum state is unavoidably coupled to its environment.

8. Statistical Aspects of Measurement

We turn now to some simple observations of a statistical nature regarding the postulates of quantum mechanics. For simplicity we deal with observables with a discrete spectrum. Let us denote the probability of measuring $A = a_n$ on a normalized, pure state $|\psi\rangle$ by f_n , so that

$$f_n = \langle \psi | P_n | \psi \rangle, \quad (38)$$

where P_n is the projector onto the n -th eigenspace of A . Then the average value of a , in the ordinary sense of statistics, is just the sum of the a_n 's weighted by the probabilities f_n . This quantity can, however, be expressed in terms of Hilbert space operations,

$$\langle a \rangle = \sum_n f_n a_n = \left\langle \psi \left| \sum_n a_n P_n \right| \psi \right\rangle = \langle \psi | A | \psi \rangle, \quad (39)$$

where we use Eq. (1.127). We will also write $\langle A \rangle = \langle a \rangle$ for this quantity. Similarly, the variance Δa^2 , defined in the usual way in statistics, can be expressed in terms of Hilbert space operations:

$$\Delta a^2 = \sum_n f_n a_n^2 - \left(\sum_n f_n a_n \right)^2 = \langle \psi | A_1^2 | \psi \rangle, \quad (40)$$

where

$$A_1 = A - \langle A \rangle. \quad (41)$$

We will also write $\Delta A^2 = \Delta a^2$. The proof of Eq. (40) is straightforward. But this equation has an immediate consequence. Let us ask for the conditions under which $\Delta a^2 = 0$, that is, the conditions under which measurements of a quantum mechanical system will yield a single value with 100% probability, with no dispersion. We write

$$\Delta a^2 = 0 = \langle \psi | A_1^\dagger A_1 | \psi \rangle = \langle \phi | \phi \rangle, \quad (42)$$

where $|\phi\rangle = A_1|\psi\rangle$. But by Eq. (1.27), this holds if and only if $|\phi\rangle = 0$, or,

$$A|\psi\rangle = \langle A \rangle |\psi\rangle. \quad (43)$$

In other words, a quantum measurement of an observable A produces no dispersion if and only if the state $|\psi\rangle$ is an eigenstate of A .

9. Generalized Uncertainty Principle

A related subject is that of the generalized uncertainty principle, which places lower bounds on the products of the dispersions of two observables, A and B . We first quote the result,

$$\Delta A^2 \Delta B^2 \geq \frac{1}{4} |\langle [A, B] \rangle|^2, \quad (44)$$

where expectation values are taken with respect to some state $|\psi\rangle$. We note that these dispersions refer to measurements of either A or B on an ensemble of identically prepared systems. They do not refer to successive measurements, an important topic discussed elsewhere in these notes.

To prove Eq. (44), we write

$$\begin{aligned} |\alpha\rangle &= A_1|\psi\rangle, \\ |\beta\rangle &= B_1|\psi\rangle, \end{aligned} \quad (45)$$

where

$$\begin{aligned} A_1 &= A - \langle A \rangle, \\ B_1 &= B - \langle B \rangle, \end{aligned} \quad (46)$$

and use the Schwarz inequality (1.28) in the form,

$$\langle \alpha | \alpha \rangle \langle \beta | \beta \rangle \geq |\langle \alpha | \beta \rangle|^2. \quad (47)$$

By Eq. (40), the left hand side of this is $\Delta A^2 \Delta B^2$. As for the right hand side, we have

$$\langle \alpha | \beta \rangle = \langle \psi | A_1 B_1 | \psi \rangle = \frac{1}{2} \langle \psi | [A_1, B_1] | \psi \rangle + \frac{1}{2} \langle \psi | \{A_1, B_1\} | \psi \rangle, \quad (48)$$

where we have written the product $A_1 B_1$ as one half the sum of the commutator and the anticommutator. Since A_1 and B_1 are Hermitian, the commutator is anti-Hermitian and the anticommutator is Hermitian. Therefore the first term on the right hand side of Eq. (48) is purely imaginary and

the second is purely real; the two terms on the right hand side are the real and imaginary parts of the expression on the left. Therefore when we take the absolute value squared of the left hand side, we obtain the inequality,

$$|\langle \alpha | \beta \rangle|^2 \geq \frac{1}{4} |\langle \psi | [A_1, B_1] | \psi \rangle|^2. \quad (49)$$

But since A and A_1 differ only by the c -number $\langle A \rangle$, which commutes with everything, and likewise for B and B_1 , the commutator in Eq. (49) can be replaced simply by $[A, B]$. Combining all the pieces, we obtain Eq. (44).

The best known example of this generalized uncertainty principle is the case $A = x$, $B = p$, $[A, B] = i\hbar$, and therefore (after taking the square root)

$$\Delta x \Delta p \geq \frac{\hbar}{2}. \quad (50)$$

When using the uncertainty principle in the form of Eq. (44), it is important to realize that the measurements of observables A and B are carried out independently on a given ensemble, represented by $|\psi\rangle$. That is, first one measures A on the ensemble, and computes $\langle A \rangle$ and ΔA ; then one measures B on the same ensemble, and computes $\langle B \rangle$ and ΔB . One does *not* measure A and B successively on individual systems in the ensemble. We talked about successive measurements in Sec. 4 above, but that is not what we are talking about when using Eq. (44).

Problems

1. Let A and B be observables, for simplicity with discrete spectra.

(a) Show that $[A, B] = 0$ if and only if $[P_{A_n}, P_{B_m}] = 0$ for all n, m , where P_{A_n} and P_{B_m} are projectors defined below Eq. (25).

(b) Show that the probabilities defined in Eqs. (29) and (30) are equal for all n, m and $|\psi_0\rangle$ if and only if $[A, B] = 0$.

2. This problem arose out of the thought that maybe the last, unknown sign multiplying μ_y in Eq. (23) could be determined if we investigated tandem Stern-Gerlach measurements at all possible orientations, not just at right angles.

In this problem we imagine that we know nothing about Pauli matrices or the formalism of spin, but we do know the measurement postulates of quantum mechanics, and we do know certain experimental facts.

In particular, suppose we have a collimated beam of silver atoms from an oven which is passed through a Stern-Gerlach apparatus oriented in the $\hat{\mathbf{n}}_1$ direction. The beam splits in two, and the beam corresponding to spin pointing in the negative $\hat{\mathbf{n}}_1$ direction is thrown away, while the beam corresponding to spin pointing in the positive $\hat{\mathbf{n}}_1$ direction is passed onto a second Stern-Gerlach

apparatus, oriented in the $\hat{\mathbf{n}}_2$ direction. Experimentally (or perhaps we should say, gedanken-experimentally), it is found that the probability that an atom entering the second apparatus will be measured to point in the positive $\hat{\mathbf{n}}_2$ direction is given by

$$\frac{1}{2}(1 + \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2) = \cos^2 \gamma/2, \quad (51)$$

where γ is the angle between $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$,

$$\cos \gamma = \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2. \quad (52)$$

We take γ to lie in the range $0 \leq \gamma \leq \pi$. The probability of measuring the spin to lie in the negative $\hat{\mathbf{n}}_2$ direction is complementary to (1), namely,

$$\frac{1}{2}(1 - \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2) = \sin^2 \gamma/2. \quad (53)$$

We notice that all spin measurements produce only two possible results, so we take the Hilbert space of our quantum system to be 2-dimensional, so that the eigenspaces of the operators which corresponds to measurements of spin are nondegenerate. We let $S(\hat{\mathbf{n}})$ or $S(\theta, \phi)$ be the operator corresponding to measuring spin with an apparatus pointing in the $\hat{\mathbf{n}} = (\theta, \phi)$ direction, and we let $|S(\hat{\mathbf{n}})\pm\rangle$ or $|S(\theta, \phi)\pm\rangle$ be normalized eigenkets corresponding to the component of spin in the $\hat{\mathbf{n}}$ direction taking on the values $\pm\hbar/2$. These eigenkets are subject to phase conventions which we will have to specify further.

(a) At first we are interested in expressing the kets $|S(\theta, \phi)\pm\rangle$ in terms of some standard eigenkets of the operator $S_z = S(\hat{\mathbf{z}}) = S(0, 0)$, which we call $|+\rangle$ and $|-\rangle$. Therefore we consider the case that $\hat{\mathbf{n}}_1 = \hat{\mathbf{n}} = (\theta, \phi)$, and $\hat{\mathbf{n}}_2 = \hat{\mathbf{z}}$. Write the kets $|S(\theta, \phi)\pm\rangle$ as linear combinations of the kets $|\pm\rangle$ with unknown coefficients. Show that by using the experimental results (51) and (53), and phase conventions for the kets $|S(\theta, \phi)\pm\rangle$, it is possible to make the coefficient of $|+\rangle$ real and nonnegative in the expression for $|S(\theta, \phi)+\rangle$, and the coefficient of $|-\rangle$ real and nonnegative in the expression for $|S(\theta, \phi)-\rangle$, and to determine all coefficients apart from a single phase factor, say, $e^{i\beta}$, where β is an as-yet-unknown function of (θ, ϕ) . Show that with these phase conventions, we have $|S(0, 0)+\rangle = |+\rangle$, and $|S(0, 0)-\rangle = |-\rangle$.

(b) Now write out the operators $S(\theta, \phi)$ in terms of the basis operators $|+\rangle\langle+|$, $|+\rangle\langle-|$, etc., and in terms of the unknown phase β . Show that by changing the phase conventions for all kets $|S(\theta, \phi)+\rangle$, including $|+\rangle$, by a constant phase factor, independent of (θ, ϕ) , and similarly by changing all kets $|S(\theta, \phi)-\rangle$, including $|-\rangle$, by a (possibly different) constant phase factor, it is possible to make $\beta(\theta, \phi)$ vanish at one chosen value of (θ, ϕ) . Do this for $(\theta, \phi) = (\pi/2, 0)$, so that $S_x = S(\pi/2, 0)$ will have real matrix elements in the $|\pm\rangle$ basis.

(c) Now return to the case of arbitrary $\hat{\mathbf{n}}_1 = (\theta_1, \phi_1)$ and $\hat{\mathbf{n}}_2 = (\theta_2, \phi_2)$, and try to use the experimental data (51) and (53) to determine the function $\beta(\theta, \phi)$. Is this function completely determined by the experimental data, or is there still some ambiguity?

Finally, prove the formula,

$$S(\hat{\mathbf{n}}) = \hat{\mathbf{n}} \cdot \mathbf{S}. \quad (54)$$

(d) Were you able to find the last, unknown sign in μ_y in Eq. (23)? If not, how do you think Pauli knew what sign to choose?

3. In quantum mechanics the measurement of one observable introduces an uncontrollable and unpredictable disturbance in the value of any observable that does not commute with the one being measured, as discussed in Sec. 6. For example, a measurement of μ_z in a Stern-Gerlach apparatus causes the values of μ_x and μ_y to become completely undetermined.

We can understand how this disturbance comes about in a classical model. To measure μ_z , we must use a magnetic field in the z -direction, but this causes μ_x and μ_y to precess, so their values on emerging from the Stern-Gerlach apparatus are different from their values when they entered. If classical mechanics were valid, we could calculate the precession angle for any particular particle in the beam, and compensate for the evolution in μ_x and μ_y . The beam has some spatial extent, however, so particles do not follow the same trajectory and their spins do not precess by the same angle, but if the size of the beam is made small enough the spread in these angles can be made as small as we like. In particular, it can be made $\ll 2\pi$, giving us a definite phase angle for the whole beam, and therefore known (and controllable) effects on μ_x and μ_y when we measure μ_z .

Show that if we try to do this in quantum mechanics, we wash out the effect we are trying to observe, namely, the splitting of the beam into two beams that make two spots on a screen. For simplicity you may assume that Eq. (37) is valid.

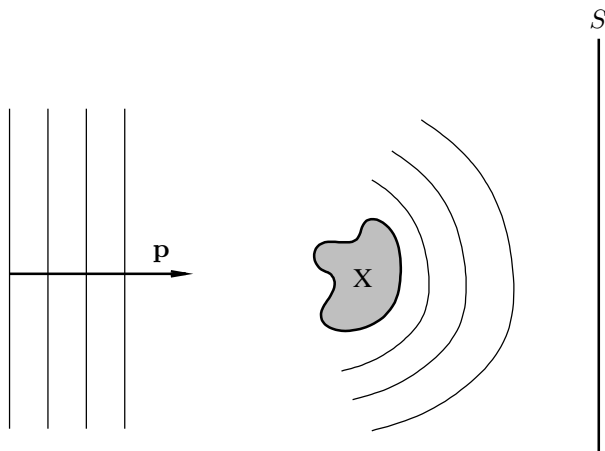


Figure for problem 4. A scattering experiment.

4. Measuring the Schrödinger wave function is not like measuring a classical field, such as an electric field. Consider a scattering gedankenexperiment in which spinless particles of a given energy are

directed at a target, as in the figure. We wish to measure the wave function downstream from the scatter. We assume the beam is described by a pure state, and that the incident wave is a plane wave (over a sufficiently large spatial region). The beam is low density, so the particles do not interact with one another. To measure $|\psi(\mathbf{r})|^2$ over some volume of space, we just put a screen S in a certain location, and gather enough statistics to get the probability density on this surface. We then move the surface and measure again.

Describe a modification to this gedankenexperiment by which the phase of the wave function $\psi(\mathbf{r})$ can be measured on the screen, apart from the overall phase, which is nonphysical and can never be measured.