The formalism of quantum mechanics

Relevant sections in text: Chapter 3

Now for some formalism...

We have now acquired a little familiarity with how quantum mechanics works via some simple examples. The approach has been rather informal; I have not systematicallty stated the rules of the game. The situation is analogous to one way of studying Newtonian mechanics. In your first exposure to Newtonian mechanics you may have had the following experience. First, you learn some basic concepts, such as vectors, force, acceleration, energy, etc. You learn, to some extent, how to use these concepts by studying some simple examples. Along the way, you identify (perhaps without full generality) some of the key laws that are being used. Finally, the whole framework is presented more systematically, *e.g.,* by carefully stating Newton's laws. More or less, we have completed the first of the steps. Now we must complete the second step.

To formulate the laws (or postulates) of quantum mechanics, we need to spend a little time giving some mathematical method to the madness we have been indulging in so far. In particular, we need to give a linear algebraic type of interpretation to the apparatus of wave functions, operators, expectation values, and so forth. We begin by developing a point of view in which states of the system are elements of a (very large) vector space. I assume that you are already familiar with the fundamentals of linear algebra (vector spaces, bases, linear transformations, eigenvalue problems, *etc.*). If you need some remedial work, have a look at the first section of chapter 3 in the text. I will only introduce the concepts that we will actually need to explain the rules of quantum mechanics.

Function spaces as infinite dimensional vector spaces

I want to view states of a quantum system as elements of a *vector space*. Recall that a vector space $\mathbf{V}$ is a set of *vectors**

$$\mathbf{V} = \{|\alpha\rangle, |\beta\rangle, \ldots\},$$

along with a set of *scalars*, $(a, b, c, \ldots)$. There is a rule for addition of any two vectors:

$$|\alpha\rangle + |\beta\rangle = |\gamma\rangle,$$

* We are using Dirac's "ket" notation for identifying vectors. We could just as well use the usual vector arrow, or bold-face, or something like that. But the ket notation has distinct advantages, as we shall see.

such that the "sum" of any two vectors is another vector. There is a rule for multiplication of vectors by scalars:

$$a|\alpha\rangle = |\beta\rangle,$$

such that the "product" of a scalar and a vector is another vector. The addition rule is assumed commutative, associative, and the scalar multiplication is distributive over addition. Scalar multiplication is also associative. Finally, there is a *zero vector*, which is an additive identity.

Simple examples of vector spaces are (1) space of position vectors with addition defined by the parallelogram rule, (2) space of n-tuples of numbers (or column vectors with n entries), with addition defined component by component. As a great exercise, verify all the above properties of a vector space in the context of these examples.

The claim is now that the space of wave functions can be viewed as forming a vector space – the vector space of states. The set is the collection of all square integrable, complex valued functions ψ of x. Here "square integrable" just means

$$\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx < \infty.$$

Certainly every normalized wave function is square integrable, and every square integrable function (except $\psi = 0$ – see below) can be normalized, so we are including all wave functions in our space. Since the vectors in our space are the wave functions we are making the notational identification:

$$|\psi\rangle \longleftrightarrow \psi(x).$$

Next, define the sum of two vectors (functions) to be the ordinary (pointwise) sum of the functions. For this to make sense we have to check that the sum of two normalizable functions is normalizable. This is not obvious, but is nevertheless true. The proof can be found in analysis and/or advanced calculus texts.

The scalars will be the complex numbers. Define scalar multiplication to be the usual multiplication of a complex function by a complex number to get another complex function. It is clear that scalar multiplication takes a square-integrable function to another square-integrable function, so this definition makes sense. You can check the usual addition of functions and multiplication of functions by scalars have all the required properties listed above.

The zero vector is just the function that assigns the value zero to every x, *i.e.,* the function $\psi(x) = 0$. Since the zero vector is not normalizable, we cannot define the vector space to be the vector space of normalized wave functions without giving up the very

convenient function $\psi(x) = 0$. For this reason I enlarged the vector space description to include all square-integrable functions.

Occasionally, additional requirements are imposed upon on our allowed wave functions. For example, the particle in a box is modeled by a wave function that vanishes outside the box. Such restrictions will not (or should not!) interfere with the vector space structure just described. Consider the particle in a box wave functions that vanish outside the open interval $x \in (0, a)$. If you add two such wave functions with this property, the result also has this property. If you multiply by a scalar, you preserve this property, the zero function has this property, and so forth.

You know that every vector space admits a *basis*. A basis is a subset of vectors, $|e_i\rangle, i = 1, 2, \ldots$, such that any vector $|\psi\rangle$ can be uniquely written as

$$|\psi\rangle = c_1|e_1\rangle + c_2|e_2\rangle + \ldots = \sum_i c_i|e_i\rangle.$$

Here the coefficients c_i are scalars. The number of basis elements is called the *dimension* of the vector space. Every finite dimensional vector space admits infinitely many bases. Our vector space of wave functions admits bases as well. For example, the stationary state wave functions for the harmonic oscillator form a basis: every square-integrable function of x can be expressed as a superposition of harmonic oscillatory stationary states:

$$\psi(x) = \sum_{n=0}^{\infty} c_n \psi_n(x).$$

As another example, the vector space of energy eigenfunctions for a particle in a box (those sine functions) form a basis. Quite generally, the stationary states (or energy eigenfunctions) for a given physical system* will form a basis for the vector space of wave functions of that system. Evidently, the vector space of wave functions is *infinite dimensional*. This is a fact of life when you work with vector spaces that are spaces of functions, *i.e., function spaces*.

Inner products

Measurements of observables generally result in real numbers. So, whenever we use vector spaces to model the world around us, we must have a way of extracting numbers from the vectors. Usually, a vector space comes equipped with an extra bit of structure called an *inner product* (also called *scalar product*). In general, an inner product is a way of assigning a scalar to any two vectors. The rule for building the scalar must be separately linear in each variable.

* Recall that stationary states or energy eigenfunctions will require a time-independent potential energy function.

In our fancy vector space notation we represent the scalar associated to vectors $|\alpha\rangle$ and $|\beta\rangle$ by $\langle\alpha|\beta\rangle$, where the following properties are postulated

$$\langle\alpha|\beta\rangle = \langle\beta|\alpha\rangle^*,$$

$$\langle\alpha|(c|\beta\rangle + d|\gamma\rangle) = c\langle\alpha|\beta\rangle + d\langle\alpha|\gamma\rangle.$$

$$\langle\alpha|\alpha\rangle \geq 0, \quad \langle\alpha|\alpha\rangle = 0 \Leftrightarrow |\alpha\rangle = 0.$$

For vector spaces defined using real scalars only the inner product does not depend upon the position of the vectors, $\langle\alpha|\beta\rangle = \langle\beta|\alpha\rangle$ (think about the dot product). In the context where scalars can be complex, one uses a scalar product which *does* depend upon the ordering of the vectors. The two different possible orderings differ by a complex conjugation. The behavior of the inner product under complex conjugation, as postulated above, is needed to guarantee that the inner product of a vector with itself is a real number. Also notice that these properties imply that, in the complex case, the inner product is *conjugate linear* in the left-most vector. What I mean is,

$$|\alpha\rangle = a|\beta\rangle \quad \Longrightarrow \quad \langle\alpha|\gamma\rangle = a^*\langle\beta|\gamma\rangle.$$

Let me mention some examples. If the vector space is the space of position vectors, then an example of an inner product is the standard dot product (exercise). If the vector space is the space of column vectors v over the set of real scalars, then an inner product of v and w is $v^t w$, where the superscript "t" means transpose. If the scalars can be complex, so that we can have complex entries in the column vector, and we denote

$$v \leftrightarrow |v\rangle, \quad w \leftrightarrow |w\rangle,$$

then

$$\langle v|w\rangle = v^{t*}w \equiv v^\dagger w.$$

You should check that this definition of inner product does satisfy the properties listed above. Note that I have used the notation †, which is called *Hermitian conjugation*. For matrices, Hermitian conjugation means composition of the operations of complex conjugation and transposition of matrices.

In quantum mechanics, with the vector space being the space of square-integrable, complex-valued functions, we can define an inner product by

$$\langle\psi|\phi\rangle = \int_{-\infty}^{\infty} \psi^*(x)\phi(x)\,dx.$$

It is a nice exercise to check that this definition satisfies all the properties of an inner product. The vector space of square-integrable functions equipped with the inner product as above is called *Hilbert space*.

Given an inner product, we can now define the length, or *norm* $||\alpha||$ of a vector $|\alpha\rangle$ to be

$$||\alpha|| = \sqrt{\langle\alpha|\alpha\rangle}.$$

We can also define two vectors $|\alpha\rangle$ and $|\beta\rangle$ to be *orthogonal* if

$$\langle\alpha|\beta\rangle = 0.$$

Given a basis $|e_i\rangle$, $i = 1, 2, \ldots$ for a vector space, we say that the basis is an *orthonormal basis* if

$$\langle e_i|e_j\rangle = \delta_{ij}.$$

The bound state wave functions (equivalently, energy eigenfunctions) for an infinite square well and for the harmonic oscillator each form an orthonormal basis for the Hilbert space of square integrable functions.

Linear Operators

An *operator* $\hat{T}$ on a vector space is just a rule that assigns to any input vector an output vector, denoted by $\hat{T}|\psi\rangle$. In certain situations (below) it will be instructive to use a slightly different notation,

$$\hat{T}|\psi\rangle \equiv |\hat{T}\psi\rangle,$$

emphasizing the fact that the operation T results in a new vector.

We call $\hat{T}$ a *linear operator* if the output vector is a *linear* function of the input vector. This means for any two vectors $\alpha\rangle$ and $|\beta\rangle$ and any two scalars a and b:

$$\hat{T}(a|\alpha\rangle + b|\beta\rangle) = a\hat{T}|\alpha\rangle + b\hat{T}|\beta\rangle.$$

The example of linear operators with which you are probably most familiar arises for the n-dimensional vector space of column vectors with n rows. The linear operators are represented by the set of $n \times n$ matrices acting on the column vectors via the usual matrix multiplication.

If the vector space is the Hilbert space of square-integrable functions, such as we use in quantum mechanics, then a linear operator T is a linear method of making a new square-integrable function $T\psi$ from any given function ψ. For example, for a given scalar a, the operator defined by

$$\hat{T}\psi(x) \equiv a\psi(x),$$

is a very simple linear operator.* The operators defined by

$$\hat{T}\psi(x) \equiv \psi(x) + 1, \quad \hat{W}\psi(x) \equiv \psi^2(x)$$

* Do not confuse this with an eigenvalue equation. This equation just says that the operator is defined as $\hat{T} = a\hat{1}$.

are *not* linear operators. More interesting examples are the *position operator*

$$\hat{x}\psi(x) = x\psi(x),$$

and the *momentum operator*

$$\hat{p}\psi(x) = \frac{\hbar}{i}\frac{d}{dx}\psi(x).$$

These are easily checked to be linear. As another example, consider the linear operator obtained by applying the momentum operator twice. We define

$$p^2\psi(x) \equiv p(p\psi(x)) = -\hbar^2\frac{d^2}{dx^2}\psi(x).$$

You can check that this operator is indeed linear. More generally, you can easily prove that the composition of two linear operators is again a linear operator.

Operator domain issues

Having defined the position and momentum operators, we must pause for a technical subtlety peculiar to infinite dimensional vector spaces. It can happen that a linear operation does not take every element of the vector space and output another element of the vector space. For example, the following function is square-integrable (exercise)

$$\psi(x) = \frac{i}{\sqrt{1+x^2}}.$$

However, if we act on it with the position operator we get the function

$$\hat{x}\psi(x) = \frac{ix}{\sqrt{1+x^2}},$$

which is *not* square-integrable (exercise). To confuse matters more, the momentum operator doesn't cause this problem with this function, since

$$\hat{p}\psi(x) = -\frac{\hbar x}{(1+x^2)^{3/2}}$$

is square-integrable. On the other hand, there are functions that are square-integrable whose derivative is not. For example, the square pulse function

$$\psi = \begin{cases} 1 & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

is square-integrable. But its derivative is a sum of two delta-functions, which is definitely not square-integrable.

The point here is that linear operators may have a restricted *domain* of definition. Not all operators have this problem. For example, T, defined by

$$\hat{T}\psi(x) = c\psi(x),$$

where c is any constant, is defined on all square-integrable functions. This domain issue can be a source of trouble, but we shall try to steer clear of it. For now, we simply point out that we have to assume that the wave function being operated on is in the domain of the operator.

Matrix elements, Hermitian operators

Given a vector space with an inner product, and given a linear operator $\hat{T}$ along with a couple of vectors, $|\alpha\rangle$ and $|\beta\rangle$, we can define complex numbers by taking the inner product of a vector $|\alpha\rangle$ with the the vector $\hat{T}|\beta\rangle$:

$$(\langle\alpha|)(\hat{T}|\beta\rangle) \equiv \langle\alpha|\hat{T}|\beta\rangle \equiv \langle\beta|\hat{T}\alpha\rangle.$$

This is called the α-β *matrix element* of $\hat{T}$. This terminology will become clear to you if you consider the case where the vector space is the space of column vectors, T is a matrix, and α and β are a couple of basis vectors. Then you can easily see that by varying the choice of the basis vectors, the indicated inner products yield all the matrix elements. You really have to try this yourself to see what I mean.

Given a linear operator $\hat{T}$, suppose we define a new linear operator, denoted by $T^\dagger$, by the requirement that

$$\langle\beta|\hat{T}\alpha\rangle = \langle\hat{T}^\dagger\beta|\alpha\rangle.$$

Note that we can also write this relation as

$$\langle\alpha|\hat{T}^\dagger|\beta\rangle = \langle\beta|\hat{T}|\alpha\rangle^*.$$

This new operator, $T^\dagger$, is called the *Hermitian adjoint* or just *adjoint* of $\hat{T}$. In our matrix example, the adjoint of a matrix is the matrix obtained by taking the complex conjugate of each entry and then taking the matrix transpose (exercise). If the operators T and $T^\dagger$ have the same domain, and on that domain $T = T^\dagger$, we say the operator T is *self-adjoint*. If we are a little more careless, and forget about domains and have the identity $T = T^\dagger$ (on some set of vectors), then we say the operator is *Hermitian*. In what follows we will follow the usual physicist's tradition and assume that Hermitian operators are also self-adjoint. (But, be warned: mathematicians will be uncomfortable with this; in general one has to verify whether a Hermitian operator is also self-adjoint.)

A simple example of a all this is as follows. Consider a 2-d complex vector space of column vectors,

$$|v\rangle = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \vec{v}.$$

A linear operator is a matrix

$$\hat{T} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \hat{T}|v\rangle = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} av_1 + bv_2 \\ cv_1 + dv_2 \end{pmatrix}$$

The Hermitian adjoint of this matrix is the complex conjugate-transpose:

$$T^\dagger = T^{*t} = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix},$$

as you can see from

$$\langle w|\hat{T}v\rangle = \vec{w}^{*t}\hat{T}\vec{v} = (\hat{T}^{*t}\vec{w})^{*t}\vec{v} = \langle \hat{T}^{*t}w|v\rangle.$$

For the matrix to be Hermitian we must have (exercise)

$$a = a^*, b = c^*, d = d^*,$$

that is

$$T = \begin{pmatrix} a & b \\ b^* & d \end{pmatrix},$$

where a and d are real numbers. (Domain issues do not arise for finite-dimensional vector spaces, so Hermitian is really synonymous with self-adjoint here.) This says the matrix elements of the complex-conjugate-transpose matrix equal the original matrix. This is consistent with the matrix element relation shown earlier:

$$\langle \alpha|\hat{T}^\dagger|\beta\rangle = \langle \beta|\hat{T}|\alpha\rangle^*.$$

Hermitian operators are easy to find in the vector space of square-integrable functions. The position operator is Hermitian:

$$\begin{aligned} \langle \phi|\hat{x}\psi\rangle &= \int_{-\infty}^{\infty} \phi^* x\psi \, dx \\ &= \int_{-\infty}^{\infty} (x\phi)^* \psi \, dx \\ &= \langle \hat{x}\phi|\psi\rangle. \end{aligned}$$

Similarly, the momentum operator is Hermitian:

$$\begin{aligned} \langle \phi|\hat{p}\psi\rangle &= \int_{-\infty}^{\infty} \phi^* \frac{\hbar}{i}\frac{d}{dx}\psi \, dx \\ &= -\int_{-\infty}^{\infty} \left(\frac{\hbar}{i}\frac{d}{dx}\phi^*\right) \psi \, dx \\ &= \int_{-\infty}^{\infty} \left(\frac{\hbar}{i}\frac{d}{dx}\phi\right)^* \psi \, dx \\ &= \langle \hat{p}\phi|\psi\rangle. \end{aligned}$$

To get the second equality I integrated by parts and threw away the boundary term since square-integrable functions vanish at infinity.

Hermitian operators are used for representing observables. There are a couple of reasons for this. The first reason manifests itself in the fact that the expectation value of a Hermitian operator is real. To see this, first note that the expectation value of an observable T in a state represented by the wave function $|\psi\rangle$ is the corresponding diagonal matrix element of its operator representative $\hat{T}$:

$$\langle T\rangle = \int_{-\infty}^{\infty} \psi^* \hat{T}\psi \, dx = \langle\psi|\hat{T}|\psi\rangle.$$

This matrix element will be real if $\hat{T}$ is Hermitian:

$$\langle T\rangle = \langle\psi|\hat{T}\psi\rangle = \langle\hat{T}\psi|\psi\rangle = \langle\psi|T\psi\rangle^* = \langle T\rangle^*.$$

To understand this better, and to see what is the second reason we use Hermitian operators to represent observables, we need to review the notions of eigenvalues and eigenvectors.

Eigenvalues and eigenvectors

Generally, a linear operator transforms vectors into other vectors. It is possible that a given operator will be such that some special vectors only get rescaled by the linear transformation:*

$$\hat{T}|\psi\rangle = \lambda|\psi\rangle.$$

When this happens we say that $|\psi\rangle$ is an *eigenvector* of T with *eigenvalue* λ. Usually these special vectors get denoted by $|\lambda\rangle$, so we can more easily recall this property.

You have already seen several examples of this. In each of the particle in the box, harmonic oscillator, and square well examples, we found solutions $|\psi\rangle$, E of the equation

$$\hat{H}|\psi\rangle = E|\psi\rangle.$$

Here the vectors $|\psi\rangle$, the energy eigenfunctions (also representing the stationary states), are the eigenvectors of the linear operator $\hat{H}$, where

$$\hat{H}\psi = -\frac{\hbar^2}{2m}\psi'' + V\psi,$$

and the corresponding energy E is the eigenvalue. Evidently, a better notation for these vectors would be $|E\rangle$, and we shall use this notation for energy eigenvectors:

$$\hat{H}|E\rangle = E|E\rangle.$$

* Here we are not defining the operator as $\lambda\hat{1}$, since that would have $\hat{T}|\psi\rangle = \lambda|\psi\rangle$ for all vectors $|\psi\rangle$. Here we are supposing that (at least) one vector and scalar satisfy this relation, but not all vectors.

The free particle example and the scattering states of the square well example led to similar relations, but since the would-be eigenfunctions are not normalizable, they are not in the vector space, strictly speaking. More on this later.

In each of our examples (particle in a box, harmonic oscillator, *etc.*) the eigenvalues E were real numbers. This is good since the eigenvalues were interpreted as the allowed energies of the system. Now, it is not too hard to see (using a combination of our proofs for position and momentum) that the Hamiltonian $\hat{H}$ is a Hermitian operator. Hermitian operators are nice since they *always* have real eigenvalues. How to see this? Just deduce it from the fact that the expectation value of the Hamiltonian in a eigenstate with energy E – denoted now by $|E\rangle$ – is real:

$$\begin{aligned}\langle H\rangle &= \langle E|\hat{H}|E\rangle \\ &= \langle E|E|E\rangle \\ &= E\langle E|E\rangle.\end{aligned}$$

Now, since the expectation value is real, and the squared-norm of $|E\rangle$ is real, it follows that E is real. Note we only used the fact that the operator was Hermitian to get the proof. Thus eigenvalues of Hermitian operators are always real.

In general, the fact that the eigenvalues E are real is in agreement with the fact that the expectation value is real (for any vector, not just eigenvectors) since the expectation value is a weighted average over the allowed energy eigenvalues.

It is easy to show that the eigenvectors of a Hermitian operator are orthogonal if they correspond to different eigenvalues. Let $\hat{T}$ be a Hermitian operator, and let $|\lambda_1\rangle$ and $|\lambda_2\rangle$ be eigenvectors with eigenvalues λ_1 and λ_2, respectively. Consider the matrix element

$$\langle\lambda_2|\hat{T}\lambda_1\rangle = \lambda_1\langle\lambda_2|\lambda_1\rangle.$$

By definition of adjoint we have

$$\langle\lambda_2|\hat{T}\lambda_1\rangle = \langle\hat{T}^\dagger\lambda_2|\hat{\lambda}_1\rangle.$$

Since T is Hermitian, we have $T^\dagger = T$ so that

$$\lambda_1\langle\lambda_2|\lambda_1\rangle = \langle\lambda_2|\hat{T}\lambda_1\rangle = \langle\hat{T}^\dagger\lambda_2|\hat{\lambda}_1\rangle = \lambda_2\langle\lambda_2|\lambda_1\rangle.$$

We conclude that

$$(\lambda_2 - \lambda_1)\langle\lambda_2|\lambda_1\rangle = 0.$$

So, provided $\lambda_1 \neq \lambda_2$, the eigenvectors are orthogonal.

Let us summarize the key features of Hermitian operators found thus far: (1) The eigenvalues are real; (2) Eigenvectors with distinct eigenvalues are orthogonal.

There is a third property of Hermitian (better: self-adjoint) operators that is extremely important. For a special class of vector spaces, the so-called *Hilbert spaces*, it can be shown that the eigenvectors of a Hermitian (really, self-adjoint) operator always form a basis for the vector space. This basis can be chosen to be orthonormal.

What is a Hilbert space? Well, we won't try to define it in detail (see the text for more information) since we won't need this much technology. Suffice it to say that a Hilbert space is a vector space with an inner product and which possesses a property known as *completeness*. This latter property is needed to get our result that the eigenvectors of Hermitian operators form a basis.

The restriction to Hilbert spaces is only needed for infinite dimensional spaces. The usual spaces found in matrix problems are finite dimensional and are *always* Hilbert spaces.

You have already seen that the stationary states, *i.e.,* eigenvectors of the Hermitian operator H, form a basis. This suggests that the space of square-integrable functions is a Hilbert space, and it is. Why this mathematical result is so useful for setting up a physical theory will be seen shortly, but note that we have already used it to our advantage in writing the general solution of the Schrödinger equation.

The spectrum

Given a linear operator T, the set of all its eigenvalues is called the *spectrum* of T. For example, the spectrum of the operator

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2$$

is the set

$$\frac{1}{2}\hbar\omega,\ \frac{3}{2}\hbar\omega,\ \frac{5}{2}\hbar\omega,\ \ldots,$$

that is,

$$E_n = (n + \frac{1}{2})\hbar\omega, \quad n = 0, 1, 2, \ldots.$$

Because the spectrum consists of a discrete set of numbers, we say that the Hamiltonian for a harmonic oscillator has a *discrete spectrum*.

It is possible that some operators will have eigenvalues that form a continuum, *i.e.,* form a *continuous spectrum*. The eigenvalue problem for such operators is a bit more subtle. A simple example is the momentum operator. Let us solve for its eigenvalues:

$$p\psi = \lambda\psi \longrightarrow \frac{d\psi}{dx} = \frac{i}{\hbar}\lambda\psi.$$

This equation is easily solved, for *any* (even complex) choice of λ, and any scalar A, we have (exercise)

$$\psi(x) = Ae^{\frac{i}{\hbar}\lambda x}.$$

You might wonder how the eigenvalue can be complex since we showed that p is Hermitian, and we showed that Hermitian operators always have real eigenvalues. The loophole is that the functions $Ae^{\frac{i}{\hbar}\lambda x}$ are not square-integrable, that is, are not elements of the vector space!

This is a general result: Hermitian operators with discrete spectrum will have eigenvectors lying in the vector space. Hermitian operators with continuous spectrum will have "generalized eigenvectors", that is, eigenvectors that live elsewhere. ("Elsewhere" can be given a careful definition using the notion of a *distribution*. We shall avoid developing the mathematics of distributions, but see below for another example of one.)

You can check the result just stated using the Hamiltonians that arise in our previous examples. For the particle in the box, the energy spectrum is discrete, and the stationary states are normalizable. For the harmonic oscillator, the energy spectrum is discrete, and the stationary states are normalizable. For the free particle, the energy spectrum is continuous and the (would be) stationary states are not normalizable. For the square well the bound state energy spectrum ($E < 0$) is discrete and the stationary states are normalizable; the scattering energy spectrum ($E > 0$) is continuous and the stationary states are not normalizable. This last example shows that an operator can have a spectrum that has a discrete part and a continuous part. This is typical when the potential energy function admits both bound states and scattering states.

Another example of an operator with a continuous spectrum is the position operator. The eigenvalue problem here is tricky. We want a function $\psi(x)$ and a constant λ such that

$$x\psi(x) = \lambda\psi(x), \quad \forall x.$$

A moment's thought will convince you that no (non-vanishing) function could possibly satisfy such an equation. Something more exotic must appear in the role of the eigenfunction. The solution turns out to be, for any choice of λ and constant A,

$$\psi(x) = A\delta(x - \lambda).$$

Here we use the Dirac "delta function", which can be viewed as the limit of a normalized Gaussian, centered at λ in the limit as the width of the Gaussian goes to zero. Roughly speaking, the delta function is a Gaussian of infinite height and zero width! Clearly, this type of limit does not yield a function in the ordinary sense, but rather defines a generalized kind of function called a distribution. Such quantities cannot be integrated in the usual way, and are not square-integrable. Indeed, the eigenvalue equation above needs a special definition to even make sense of it. Here is one way – using Fourier analysis – of building a delta function satisfying the above eigenvalue equation:

$$\delta(x - \lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ik(x-\lambda)}.$$

We will not need to go into all this, but I point it out since it is a fact of life in quantum mechanics.

We have seen that operators with continuous spectrum do not behave very much like operators (matrices) on finite-dimensional spaces. In particular, Hermitian operators with continuous spectrum do not, strictly speaking, have eigenvalues or eigenvectors. Even if we allow the more general (distributional) type of eigenvectors, the eigenvalues need not be real. A more detailed examination of the issue shows that, for Hilbert spaces, (1) Hermitian operators with continuous spectrum can always have their generalized eigenvectors chosen to yield real eigenvalues; (2) the generalized eigenvectors with real eigenvalues form a sort of basis for the space in the sense that elements of the vector space *can* be represented as a superposition of the generalized eigenvectors... even if these generalized eigenvectors do not live in the vector space!

To prove this in general takes us too far afield, but an important and illustrative example arises with the momentum operator. We can choose the eigenvalues to be real; the (generalized) eigenfunction ψ_λ with eigenvalue λ can be written as

$$\psi_\lambda = \frac{1}{\sqrt{2\pi}} e^{\frac{i}{\hbar}\lambda x}.$$

We have mentioned before that, according to the theory of Fourier analysis, every square-integrable function $\psi(x)$ (*i.e.,* element of the vector space) can be written as

$$\psi(x) = \int_{-\infty}^{\infty} \phi(\lambda)\psi_\lambda(x),$$

for some choice of $\phi(\lambda)$. You should think of this as a generalized, continuous version of the result that eigenvectors of a Hermitian operator form a basis. Indeed, our last continuous superposition formula holds for any Hermitian operator where $\psi_\lambda(x)$ is the eigenfunction with continuous eigenvalue λ. If the operator has both continuous and discrete parts to its spectrum, we get a sum over the discrete eigenvalues and an integral over the continuous ones.

Postulates of quantum mechanics

We have now acquired enough terminology to succinctly state the rules of quantum mechanics. These rules provide a mathematical means of modeling the physical world in a way which, so far, has agreed with experiment in all respects.

While all of our applications of the rules of quantum mechanics will be non-relativistic (all velocities are small compared to that of light; all gravitational fields are weak) the postulates themselves are not sensitive to whether one incorporates relativity or not. Of course, relativistic effects in quantum mechanics *are* detectable, *e.g.,* in the spectrum of

the hydrogen atom, and so should be included in a more complete theory. Moreover, relativistic quantum theory sits on an uncomfortable fence between non-relativistic quantum theory and *quantum field theory*. The problem is that relativistic quantum mechanics allows for arbitrarily large energies, but high-energy processes lead to particle creation and annihilation, which quantum mechanics cannot handle in its usual implementation. One generalizes the application of the rules of quantum mechanics to fields instead of particles – quantum field theory – to accommodate such situations. Again, the postulates we are about to state still hold, and form the foundation of the quantum field theory, but the details of the implementation of the rules is rather different in quantum field theory than in quantum mechanics.

Finally, whether working in non-relativistic quantum mechanics, relativistic quantum mechanics, or quantum field theory, there is an additional feature of the world to be taken into account when considering systems with more than one particle. More precisely, one must include a way of properly handling *identical particles*. In quantum mechanics, this leads to another postulate (the symmetrization postulate), that we shall discuss a little later.

Postulate: The state of a particle is represented by a normalized vector in a Hilbert space

We have seen that the set of square-integrable functions forms a Hilbert space. Every square-integrable function (except the 0 function) can, by multiplication by a constant, be normalized (exercise). The normalization requirement is crucial for the probability interpretation, which will be stated shortly. The Hilbert space requirement is needed so that the eigenvectors of Hermitian operators form a basis, which will also be crucial for one of the postulates that is coming up.

One often says that a state of a particle (or, more generally, any quantum system) is represented by a *ray* in the vector space. In Euclidean space, a ray is a directed half-line. It is easy to see that any two non-zero vectors in Euclidean space that are parallel lie on the same ray (exercise). Moreover, a ray is the set of endpoints of vectors obtained by taking a position vector and scaling it by all positive real numbers. We can generalize this notion of ray to any vector space. A ray is the subset of vectors all of which differ by a scalar multiple. Now you can see why a ray is a state: given a ray, we simply select the element of the ray that has unit length to get the wave function. Conversely, given a wave function the ray is determined.

Postulate: Observable quantities correspond to self-adjoint operators.

Here self-adjoint means "Hermitian" with appropriate care taken with domains of

definition of operators. This postulate is needed so that (i) outcomes of measurements of observables can be represented by real numbers, and (ii) eigenvectors form a basis. The utility of (i) and (ii) is explained in the postulates that follow.

For a particle moving in one-dimension the simplest observables are (represented by) the position and momentum operators x and p. Usually, one *assumes* that no other observable features of the particle exist, that is, all other observables can be built from these two operators.* Thus, for example, the potential energy observable is $V = V(x)$. What this means is that V is represented by the Hermitian linear operator

$$V\psi(x) = V(x)\psi(x).$$

Likewise, the kinetic energy observable is represented by $K = p^2/2m$, that is,

$$K\psi(x) \equiv \frac{1}{2m}p^2\psi(x) = -\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2}.$$

It can be difficult to directly measure the kinetic or potential energy functions. It is often easier (using conservation of energy ideas) to measure the total energy, which is represented by the operator

$$H = K + V.$$

Postulate: A measurement of an observable represented by an operator T will return an element of the spectrum of T.

We see here a generalization to all observables of our previous rule that the allowed energies of the particle are the separation constants E that appear in the TISE – the spectrum of the Hamiltonian. For the particle in the box, the harmonic oscillator, and the bound states of the square well, the Hamiltonian has a discrete spectrum of allowed energies. One says that these observables have been "quantized" relative to their classical counterparts. When we come to study three-dimensional systems we will find that angular momentum is also quantized, that is, it is represented by (3) operators with discrete spectrum. Other observables, like the position, momentum, kinetic energy, scattering energies for a square well potential, *etc.* have continuous spectra and so the allowed values of a measurement of such quantities form a continuum.

Postulate: Given an observable represented by T, the probability for finding the eigenvalue λ coming from the discrete part of the spectrum of T upon

* This is certainly the case in classical mechanics. Nevertheless, it is an assumption made to simplify the model and will have to be modified eventually to take account of things like intrinsic *spin*.

measurement in the state $|\psi\rangle$ is the absolute square of the component of $|\psi\rangle$ along the basis element/eigenvector $|\lambda\rangle$.

What this postulate means is this. Suppose first that the spectrum of T is discrete, with eigenvalues λ_n, $n = 1, 2, \ldots$. Since the corresponding eigenvectors $|\lambda_n\rangle$ of T form an orthonormal basis, we can write any vector as

$$|\psi\rangle = \sum_{n=1}^{\infty} c_n |\lambda_n\rangle.$$

The probability, $P(\lambda_n)$, for getting the value λ_n in the state defined by the normalized vector $|\psi\rangle$ is given by

$$P(\lambda_n) = |c_n|^2.$$

The normalization condition implies (exercise)

$$\sum_{n=1}^{\infty} |c_n|^2 = 1.$$

Here you begin to see why it was crucial that the eigenvectors formed a basis. You can think of the complex numbers c_n as akin to a wave function that has a discrete argument (n).

As a nice exercise you can verify that this postulate implies that the expectation value of T in the state $|\psi\rangle$ is given by

$$\langle T \rangle = \langle \psi | T | \psi \rangle.$$

Another consequence of this postulate is that the probability for finding the eigenvalue λ_n (for some fixed value of n) is one if and only if the state is an *eigenstate* of T associated to the eigenvalue λ_n. What this means is that, if the state is represented by the vector $|\lambda_n\rangle$, then it is clear that all the coefficients c_i, $i \neq n$ must vanish. In this case, by normalization, $|c_n|^2 = 1$. Conversely, if the probability for finding λ_n is unity, it follows that $|c_i|^2$, $i \neq n$ must vanish, which means that $|\psi\rangle$ is an eigenvector of T with eigenvalue λ_n. We have seen this situation when dealing with stationary states (exercise); the rules of quantum mechanics assert that this situation applies to *any* observable – assuming the eigenvectors are elements of the Hilbert space. This last comment will be explored below.

Summary of the postulates so far...

So far, we have spelled out four postulates of quantum mechanics. The first two (states are normalized vectors in Hilbert space, observables are represented by Hermitian

operators) basically tell us what are the mathematical tools we should use to model a physical system. The latter two postulates (outcome of measurement is an eigenvalue, probability distribution) spell out what the physical predictions of quantum mechanics will be, given a representation of observables by operators on a Hilbert space of states. Thus you can think of the first two postulates as controlling the input to the quantum mechanical theory, and the last two as the output. Let me emphasize that this is the whole output of quantum mechanics: (1) possible outcome of measurements, and (2) probabilities for those outcomes.

The case of continuous spectra

When the spectrum of an operator representing an observable is continuous, or has a continuous part, we have to generalize slightly our discussion of the physical output of quantum mechanics. To begin, let us suppose that the spectrum has no discrete part - as in the momentum operator. Let us denote the operator by T and the eigenvalues by λ. As mentioned earlier, for Hermitian operators on Hilbert space one can restrict attention to the real part $\{\lambda\}$ of the continuous spectrum and the corresponding generalized eigenvectors $|\lambda\rangle'$ are *complete* so that, for any vector $|\psi\rangle$

$$|\psi\rangle = \int_{spectrum} \phi(\lambda)|\lambda\rangle' \, d\lambda.$$

Here I have used a prime to remind us that, strictly speaking, the generalized eigenvectors are not elements of the Hilbert space. Usually, one doesn't bother to include the prime notation. I will temporarily include the prime for pedagogical reasons.

The probability P([a,b]) for finding the observable to have a value $\lambda \in [a, b]$ is given by

$$P([a, b]) = \int_a^b |\phi(\lambda)|^2 \, d\lambda.$$

Normalization means that

$$\int_{spectrum} |\phi(\lambda)|^2 \, d\lambda = 1.$$

Letting the region $[a, b]$ be infinitesimal, we can say that $|\phi(\lambda)|^2$ is the probability density for measurement of T. This means that

$$\langle T \rangle = \langle \psi | T | \psi \rangle = \int_{spectrum} |\phi(\lambda)|^2 \lambda \, d\lambda.$$

All of this shows that the functions $\phi(\lambda)$ are used to describe the observable T, just like the position wave function is used to describe the observable x. Indeed, the position observable, its wave function, etc., should be viewed as a special case of this more general point of view.

Let us see how the position probability fits into this picture. The position eigenfunction with eigenvalue λ is given by a delta function:

$$|x = \lambda\rangle \leftrightarrow \delta(x - \lambda).$$

We can write the wave function as

$$\psi(x) = \int_{-\infty}^{\infty} d\lambda \psi(\lambda)\delta(x - \lambda).$$

Evidently, the probability density for position at $x = \lambda$ is given by $|\psi(\lambda)|^2$, in agreement with all we've done in the past.

Another nice example of this is the momentum operator. We saw that, by representing momentum with a derivative operator, the spectrum was the set of real numbers. The generalized eigenvectors are the functions

$$\psi_\lambda(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}.$$

The eigenvalue relation is then

$$p\psi = \frac{\hbar}{i}\frac{d\psi}{dx} = \hbar k\psi,$$

that is,

$$\lambda = \hbar k.$$

The numerical factor in front of the momentum eigenfunction was chosen so that the expansion of a wave function in the basis of (generalized) momentum eigenfunctions, according to the formula above, is the standard Fourier expansion:

$$\psi(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{ikx}\phi(k)\, dk.$$

The expansion coefficients $\phi(k)$ constitute the Fourier transform of $\psi(x)$. These are often called the *momentum space wave functions* since, according to our general scheme, $|\phi(k)|^2$ is the probability density for finding the momentum at the value $\hbar k$ (exercise). From this it follows that the expectation value of momentum in in the state $\psi(x)$ with Fourier transform $\phi(k)$ is given by (exercise)

$$\langle p\rangle = \int_{-\infty}^{\infty} |\phi(k)|^2 \hbar k\, dk$$

I now give our last postulate, for now. This one deals with the implementation of dynamics.

Postulate: Given a Hamiltonian, the time evolution of a given initial state is determined by the Schrödinger equation. Immediately after a measurement of an observable T with outcome given by the eigenvalue λ of $\hat{T}$, the state of the system is the corresponding eigenvector $|\lambda\rangle$.

The first part of the postulate is, by now, pretty familiar to you. To describe dynamics, one has to decide what operator represents the Hamiltonian of the system. Often times the classical approximation is a guide, but there is no universal prescription for choosing the Hamiltonian. Indeed, the art of quantum mechanics lies, at least in part, in deducing the Hamiltonian that describes the physical system of interest.

The second part of the postulate gives help in understanding how to "prepare a state". For example, suppose we want to prepare some harmonic oscillators in their ground state so as to, say, make some position measurements. We can take a box full of harmonic oscillators and measure their energies. We keep the ones with energy $\frac{1}{2}\hbar\omega$, henceforth we know these are in the desired stationary state. We do not have to prepare states only according to their energy. We can measure whichever observables we want (subject to an uncertainty relation obstruction to be discussed below) and select the systems with desired eigenvalues for these observables. Such states will, in general, change in time but they have the desired characteristics immediately after the measurement.

If the observable being measured has continuous spectrum, then the state of the system cannot be its eigenfunction, strictly speaking since the (would-be) eigenfunction is not normalizable. Instead, if our measurement localized the value of λ to be, say, in the interval (a, b), then the state of the system is a normalized superposition over generalized eigenvectors with that range of values:

$$|\psi\rangle = \frac{1}{\sqrt{b-a}} \int_a^b |\lambda\rangle \, d\lambda.$$

For example, if we make a position measurement and find $x \in (x_1, x_2)$ then the resulting state is

$$\psi(x) = \begin{cases} \frac{1}{\sqrt{x_2 - x_1}} & x_1 < x < x_2 \\ 0 & x < x_1, \ x > x_2. \end{cases}$$

As another example, suppose we measure the momentum to be in the region ($p_1 = \hbar k_1, p_2 = \hbar k_2$). Then the wave function just after the measurement is

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar(p_2 - p_1)}} \int_{p_1}^{p_2} e^{ipx/\hbar} dp = \sqrt{\frac{\hbar}{2\pi(p_2 - p_1)}} \int_{k_1}^{k_2} e^{ikx} \, dk = \sqrt{\frac{\hbar}{2\pi(k_2 - k_1)}} \frac{e^{ik_1 x} - e^{ik_2 x}}{ix}.$$

If you like taking Fourier transforms, you can check that the momentum space wave function corresponding to $\psi(x)$ (*i.e.,* the Fourier transform of $\psi(x)$) is a constant in the region

$k = (a, b)$ and zero elsewhere in "k-space". The probability density in position space is easily computed:

$$|\psi(x)|^2 = \frac{\hbar}{\pi(k_2 - k_1)} \frac{1 - \cos[(k_2 - k_1)x]}{x^2}.$$

This density is peaked about $x = 0$ and oscillates as its amplitude decreases away from $x = 0$.

This last part of the postulate is actually pretty weird. To see what I mean, suppose that a particle in a box of size 1 meter is in a superposition of the ground and first excited states at some time, say $t = 0$:

$$\Psi(x, 0) = \frac{1}{\sqrt{2}}(\psi_1(x) + \psi_2(x)).$$

If you make an energy measurement at that time you have a 50-50 chance of getting the ground state energy *versus* the first excited state energy. Suppose you get the ground state energy. If you immediately repeat the measurement you will get the ground state energy with probability one, since the state is now the ground state. Somehow, it is as if – at least mathematically – the measurement "collapsed" the wave function from the superposition of two eigenstates to the single eigenstate. This can lead to some pretty weird experimental predictions – predictions which have been verified experimentally! For example, the probability for getting the ground state energy at some time later, will depend upon whether or not you chose to measure the energy at $t = 0$. That is, if you do *not* measure the energy at $t = 0$, then you still get a 50-50 possibility of finding ground or first excited state energies when you finally do measure the energy. If someone sneaks up and secretly takes a measurement at $t = 0$, when you get around to making your measurement of energy, you will find with probability one the energy your mischievous colleague found at $t = 0$.

Another very simple example: after finding the ground state energy at $t = 0$, we then measure the position and find the particle to be, say, in the region $x = 0.24 \pm .1\, m$. Then the position wave function is, immediately after the measurement, constant in the region $x = .24 \pm .1m$ and zero elsewhere (as we worked out earlier). This is certainly not a stationary state, let alone the ground state. An immediately subsequent measurement of the energy can yield a variety of possible outcomes with varying probabilities (exercise). If we did not make the position measurement, the result of the second energy measurement would, again, be the ground state with probability one. Thus the act of measurement can affect the subsequent behavior of the system!

This last example shows that the position and energy observables are in a certain sense "incompatible", in that the measurement of one disturbs the value of the other. This is an instance of the uncertainty relation, which we shall now discuss in some detail.

The general form of the uncertainty relation for observables

I would now like to state the general form of the celebrated uncertainty relation for observables in quantum mechanics. We shall see that the uncertainty principle we have seen from time to time (*e.g.,* dealing with position and momentum observables) is a special case of this general result. I will not prove this general result! You could easily understand the proof, but I cannot improve on the text's presentation, which you can read as easily as I can. Moreover, the proof does not really help us to learn anything physically (although it nicely illustrates mathematical aspects of operators and vector spaces). I would rather spend time explaining the relation than deriving it.

To begin, it is important to emphasize that the uncertainty relation is a statement about a *pair* of observables. Moreover, the relation may depend upon the particular state of the system. To describe the uncertainty relation, we first need a definition.

Let A and B be a pair of observables, with corresponding Hermitian operators $\hat{A}$ and $\hat{B}$. Define the *commutator* $[\hat{A}, \hat{B}]$ of two operators to be the operator obtained by

$$[\hat{A}, \hat{B}]|\psi\rangle = (\hat{A}\hat{B} - \hat{B}\hat{A})|\psi\rangle.$$

The commutator is a linear operator that measures whether or not the order of the two operations matters. If the operator you get is the zero operator (maps every state to the zero vector), then the order of operation does not matter, and we say $\hat{A}$ and $\hat{B}$ *commute*. Some good exercises are to prove that:

$$[\hat{x}, \hat{x}] = 0, \quad [\hat{x}, \hat{p}] = i\hbar\hat{1}, \quad [\hat{x}^2, \hat{p}] = 2i\hbar\hat{x}, \quad [\hat{x}, \hat{p}^2] = 2i\hbar\hat{p}.$$

Here 0 is the "zero operator", which maps every vector to the zero vector; $\hat{1}$ is the identity operator, $\hat{1}|\psi\rangle = |\psi\rangle$. The square of an operator means to apply the operation twice.

Let $|\psi\rangle$ define a state, that is, $|\psi\rangle$ is a normalized element of a Hilbert space. For the observables A and B we can compute the standard deviations for the probability distributions:

$$\sigma_A = \left[\langle\psi|\hat{A}^2|\psi\rangle - \langle\psi|\hat{A}|\psi\rangle^2\right]^{1/2}$$

$$\sigma_B = \left[\langle\psi|\hat{B}^2|\psi\rangle - \langle\psi|\hat{B}|\psi\rangle^2\right]^{1/2}$$

The uncertainty relation, in its most general form, says that

$$\sigma_A\sigma_B \geq \frac{1}{2}\left|\langle\psi|[\hat{A}, \hat{B}]|\psi\rangle\right|.$$

It is proved using the *Schwarz inequality* (see your text).

The meaning of this uncertainty relation is as follows. First, the relation gives a lower bound on the product of the two standard deviations. Since, by its definition, an

uncertainty is greater than or equal to zero, if the lower bound provided by the right hand side of the uncertainty relation turns out to be zero for the given operators and the given state, then we learn nothing from the uncertainty relation. The lower bound can vanish because (i) the commutator vanishes, and/or (ii) the expectation value of the commutator vanishes in the chosen state. If this lower bound is not zero, neither of the standard deviations can vanish in the given state. A necessary condition for this is that the operators representing the observables do not commute. If they don't commute, then there is at least one state for which the right hand side of the uncertainty relation is not zero. Such observables are called *incompatible*. If the operators commute commute, the right hand side of the uncertainty relation is always zero; the corresponding observables A and B are called *compatible*. Recall that the standard deviation of an observable vanishes if and only if the state is an eigenvector of the observable. Consequently, if the uncertainty relation is non-trivial, then it implies that neither of the observables in question are known with probability one *in the chosen state*, and it gives some information on the relative sizes of the standard deviations. Depending upon the commutator and the state, this information may or may not be terribly useful.

The most striking results occur when the right hand side of the uncertainty relation is non-zero and *independent of the state*. We will see an example below where this happens. In this case, the standard deviations can never vanish, for any state. For such observables it is always impossible to establish their values with probability unity! Moreover, if one prepares a state with smaller and smaller standard deviation for one of the observables, A, say, then the standard deviation of the other observable B will, eventually, have to increase. Let us look at some examples.

The position-momentum uncertainty relation is especially dramatic since it is one where the right hand side of the relation is non-zero and independent of the state. We have

$$[\hat{x}, \hat{p}] = i\hbar\hat{1},$$

so in any state $|\psi\rangle$ (exercise),

$$\sigma_x \sigma_p \geq \frac{1}{2}|\langle\psi|\, i\hbar\hat{1}\, |\psi\rangle| = \frac{\hbar}{2}.$$

We have already explored this idea using the Fourier transform and Gaussian wave functions in position and momentum space. (Review this now!) Note in particular that the standard deviation of position and/or momentum can never vanish. This is equivalent to the fact that the "eigenfunctions" of position and momentum do not really define normalizable functions, that is, they are not in the Hilbert space.

Another incompatible observable pair is x^2 and p. We get

$$\sigma_{x^2} \sigma_p \geq \frac{1}{2}|\langle\psi|\, i\hbar\hat{x}\, |\psi\rangle| = \frac{\hbar}{2}\langle x\rangle.$$

This uncertainty relation depends upon the state. For example, suppose the state is a stationary state for the harmonic oscillator. It is not too hard to see that in any such state $\langle x \rangle = 0$. So, the uncertainty relation is trivial for stationary states. On the other hand, for a state that is a superposition of two stationary states, the expectation value of x can be non-zero, as you have seen in your homework. Thus, for such a state, neither the standard deviation of x^2 nor that of p can vanish.

To understand a little better what is going on in the uncertainty relation, and in particular why the commutator is appearing, let me remind you that an observable T is known to take the (eigen)value λ with certainty (probability one) if and only if the state of the system is an eigenvector $|\lambda\rangle$ of the corresponding operator $\hat{T}$. If two observables, A and B, are both known to have values a and b with certainty, then the state of the system $|a, b\rangle$ must be both an eigenvector of $\hat{A}$ and of $\hat{B}$:

$$\hat{A}|a, b\rangle = a|a, b\rangle, \quad \hat{B}|a, b\rangle = b|a, b\rangle.$$

This easily implies (exercise):

$$[\hat{A}, \hat{B}]|a, b\rangle = 0 \Longrightarrow \langle a, b|[\hat{A}, \hat{B}]|a, b\rangle = 0.$$

We know that eigenvectors of a Hermitian operator form a basis for Hilbert space. It is now worth noting a very important math fact: *a necessary and sufficient condition for two Hermitian operators $\hat{A}$ and $\hat{B}$ to admit a common basis of eigenvectors is that the two operators commute*. Commuting Hermitian operators represent compatible operators and their common basis of eigenvectors are the states in which the two observables are known with certainty.

Energy-time uncertainty relation

There is another kind of uncertainty relation that applies to energy and time. This relation is sometimes presented as being the same type of relation we just described for observables. While there are some superficial parallels, the time-energy uncertainty relation is physically (and mathematically) quite different. In particular, the parameter we call "time" is not treated in non-relativistic quantum mechanics as an observable of a system (such as position and momentum are for a particle). So time is not represented as a linear operator, and it does not have a probability distribution associated with a state, etc. The time-energy uncertainty relation is nonetheless quite important, if a little more qualitative, than the uncertainty relation for observables.

The basic idea of the time-energy uncertainty relation is quite simple. If a system is in a state where the energy is known with probability one, *i.e.,* a stationary state – an energy eigenstate, then it remains there forever, *i.e.,* doesn't change. If the system

is not in a stationary state, then the probability distribution for the energy observable – the Hamiltonian H – has a non-zero standard deviation σ_H. As a result, the state of the system changes in time (via the Schrödinger equation). Evidently there is some relation between σ_H and the time evolution of a state. The time-energy uncertainty relation relates the spread in energy $\sigma_H \equiv \Delta E$ to the scale of time Δt it takes for the state of the system to change appreciably from the initial state. For an appropriate choice of Δt, this relation is

$$\Delta E \Delta t \geq \frac{\hbar}{2}.$$

The key step in deriving/interpreting the energy-time uncertainty relation is to figure out how to characterize the time it takes for a state to "change appreciably". Of course, to determine what the state is doing we only have access to measurements of observables. Let us focus on the time rate of change of the expectation value of some chosen observable $Q = Q(x, p)$ in the state represented by the wave function $\Psi(x, t)$. I will now derive a fundamental formula for this time rate of change. Start with

$$\langle Q \rangle = \langle \Psi, t | \hat{Q} | \Psi, t \rangle = \int_{-\infty}^{\infty} \Psi^*(x,t) \hat{Q} \Psi(x,t)\, dx,$$

so that

$$\frac{d}{dt}\langle Q \rangle = \int_{\infty}^{\infty} \frac{\partial}{\partial t} \left(\Psi^*(x,t) \hat{Q} \Psi(x,t) \right) dx = \int_{\infty}^{\infty} \left(\frac{\partial \Psi^*}{\partial t} (\hat{Q}\Psi) + \Psi^* (\hat{Q} \frac{\partial \Psi}{\partial t}) \right) dx.$$

Now, the time evolution of the wave function is controlled by the Schrödinger equation, so we get

$$\frac{d}{dt}\langle Q \rangle = \int_{\infty}^{\infty} \left(-\frac{1}{i\hbar} (\hat{H}\Psi)^* (\hat{Q}\Psi) + \Psi^* \hat{Q} (\frac{1}{i\hbar} \hat{H} \Psi) \right) dx.$$

Because $\hat{H}$ is Hermitian, we can move it to act on the other function in the product of the first term:

$$\int_{-\infty}^{\infty} (\hat{H}\Psi)^* (\hat{Q}\Psi)\, dx = \int_{-\infty}^{\infty} (\Psi)^* (\hat{H}\hat{Q}\Psi)\, dx.$$

We then get

$$\frac{d}{dt}\langle Q \rangle = \int_{\infty}^{\infty} \Psi^* \frac{i}{\hbar} [\hat{H}, \hat{Q}] \Psi.$$

This result,

$$\frac{d}{dt}\langle Q \rangle = \langle \frac{1}{i\hbar} [\hat{Q}, \hat{H}] \rangle,$$

is quite important irrespective of the energy-time uncertainty relation. By letting $\hat{H}$ take its usual kinetic + potential form, and letting $\hat{Q}$ be position or momentum, you reproduce Ehrenfest's theorem, of which the equation above is the general form. Also, you can see that if Q is compatible with the Hamiltonian, so that $[\hat{Q}, \hat{H}] = 0$, then the expectation

value of Q is independent of time – for any state. This is a signal that Q is a conserved quantity. A trivial example of this is the energy itself. As another (almost trivial) example, let Q be momentum and H the Hamiltonian for a free particle.*

Since all predictions about observables in quantum mechanics can be (if desired) framed in terms of expectation values, the preceding formula is a fundamental expression of time evolution in quantum mechanics. Let me give you the most general version of this result by allowing the observable to explicitly depend upon time for its definition. If $Q = Q(x, p, t)$, *e.g.,* $Q = p^2 + e^t x^2$. You can easily repeat the previous analysis to find:

$$\frac{d}{dt}\langle Q\rangle = \langle \frac{1}{i\hbar}[\hat{Q}, \hat{H}] + \frac{\partial \hat{Q}}{\partial t}\rangle,$$

To get the energy-time uncertainty relation, we apply the general uncertainty relation to the observables Q and H and use our fundamental identity, just derived. Suppose that $Q = Q(x, p)$, then (exercise)

$$\sigma_H \sigma_Q \geq \frac{1}{2}|\langle [\hat{H}, \hat{Q}]\rangle| = \frac{\hbar}{2}|\frac{d\langle Q\rangle}{dt}|.$$

Define Δt by

$$\Delta t \equiv \frac{\sigma_Q}{|\frac{d\langle Q\rangle}{dt}|}.$$

You can see that Δt is an estimate for the time it takes for $\langle Q\rangle$ to change by one standard deviation. This choice of Δt is useful since, if the original statistical uncertainty of Q is σ_Q, then it hardly makes sense to say that Q has changed in a significant fashion until the change is at least as big as the statistical uncertainty you started with. If we set $\Delta E \equiv \sigma_H$, then we have (exercise)

$$\Delta E \Delta t \geq \frac{\hbar}{2}.$$

This inequality has a simple enough meaning (which is often mis-stated): If the spread of energies ΔE in the initial state is very small, then it will take a relatively long time for observables to change appreciably. Conversely, if some observable changes very rapidly then the spread of energies must be large. The precise estimate on Δt depends upon the observable and the state chosen. As an extreme limiting case, suppose the initial state is a stationary state. Then $\Delta E = 0$, and the energy-time uncertainty relation has $\Delta t \to \infty$, as you expect for a stationary state.

You can see from this discussion that the Δt in the time-energy uncertainty relation is *not* to be interpreted as some kind of statistical uncertainty in our knowledge of time, as may be inferred from popular characterizations of this relation.

* The relative scarcity of interesting conservation laws (besides energy) for systems in one dimension arises from the small number of degrees of freedom, and occurs already classically.

It is often said that the energy-time uncertainty relation means that you can violate conservation of energy by an amount ΔE for a time Δt. This is an extremely misleading statement, and unless interpreted very carefully, just plain wrong. Beware of such statements in the popular literature. For closed systems (with a time-independent Hamiltonian) the energy is always conserved. The meaning of energy conservation in quantum mechanics is twofold. First, the probability distribution of outcomes of energy measurements does not depend upon time. This is the best you can hope for if the state of the system is not one of definite energy. Second, if the state of the system is initially known to have some fixed energy with probability one, then the initial state is a stationary state with that energy, and remains so for all time. Thus if the initial energy is known to be E, then it remains E for all time.*.

Let me exhibit a simple example of the energy-time uncertainty relation. Consider a system with an initial wave function which is a superposition of two stationary states:

$$\Psi(x,0) = \frac{1}{\sqrt{2}}(\psi_1(x) + \psi_2(x)).$$

You can easily check that, denoting the corresponding energies by E_1 and E_2, we have

$$\langle H \rangle = \frac{1}{2}(E_1 + E_2), \quad \langle H^2 \rangle = \frac{1}{2}(E_1^2 + E_2^2),$$

so that

$$\Delta E = \frac{1}{\sqrt{2}}|E_1 - E_2|.$$

According to the time-energy uncertainty relation, the initial state has changed appreciably when

$$\Delta t \geq \frac{\hbar}{\sqrt{2}|E_1 - E_2|}.$$

To check this with some observable, focus on position. Consider the probability density for position at time t. Since

$$\Psi(x,t) = \frac{1}{\sqrt{2}}\left(\psi_1(x)e^{-\frac{i}{\hbar}E_1 t} + \psi_2(x)e^{-\frac{i}{\hbar}E_2 t}\right),$$

we have

$$\rho(x,t) = \frac{1}{2}\left[|\psi_1(x)|^2 + |\psi_2(x)|^2 + 2\Re\left(\psi_1(x)\psi_2^*(x)\right)\cos\left(\frac{E_1 - E_2}{\hbar}\,t\right)\right].$$

At $t = 0$ the cosine is one; as time passes the cosine decreases, changing the probability distribution for position. When

$$\Delta t = \frac{\hbar}{\sqrt{2}|E_1 - E_2|}$$

* All these statements are about time evolution *between measurements*.

the cosine term becomes $\cos(\sqrt{2}/2) \approx 0.76$.

Another example of the energy-time uncertainty relation is as follows. Consider a free particle wave packet with standard deviation Δx. We have seen that the peak of the probability distribution in position space moves with some fixed (group) velocity v. How long does it take for the wave packet to pass a given point? We can give an estimate of this time by finding the time it takes for the expectation value of position to change by one standard deviation. The packet moves with speed v, so the time is

$$\Delta t = \frac{\Delta x}{v}.$$

To check the energy-time uncertainty relation. Recall that the expectation value of momentum is $\langle p \rangle = mv$; the expectation value of energy is $\langle E \rangle = \frac{\langle p^2 \rangle}{2m}$. The energy, like the momentum, has a standard deviation, which we can estimate as (exercise):

$$\Delta E = \frac{\langle p \rangle}{m} \Delta p.$$

Of course, all of this can be worked out more precisely, but the estimates are good enough for our discussion. Given our estimates, the uncertainty product is

$$\Delta E \Delta t = \frac{\langle p \rangle}{m} \Delta p \frac{\Delta x}{v} = \Delta x \Delta p \geq \frac{\hbar}{2},$$

which is consistent with our general result.

Another way the energy-time uncertainty relation is used is to describe decay processes. Suppose a particle is known to decay with a half life of Δt. If you measure the rest energy mc^2 of the particle you will find a statistical distribution with standard deviation ΔE. The relation between ΔE and Δt will be found to be

$$\Delta E \Delta t \approx \hbar \geq \frac{\hbar}{2}.$$

Thus the longer the lifetime of a particle, the more well defined is its rest energy. A similar effect is observed when measuring energy levels of atoms. As you know, because of the coupling of the atom to the electromagnetic field, the excited states are not stable, but will in general decay to lower energies by emitting one or more photons. The lifetime of a given energy level depends upon the detailed properties of that state. What is observed is that the measured energy of a given level has a certain intrinsic spread ΔE (standard deviation), which is inversely proportional to the half-life Δt of the state, such that

$$\Delta E \Delta t \approx \hbar.$$

The more long-lived is the excited state, the more precisely defined is its energy.

You can see that the energy-time uncertainty relation can be used in a variety of ways. The various applications depend upon what observable you are using to characterize Δt.