Quantum Mechanics

or

Watching a Theoretical Physicist at Work

Chapter 1

Motivation, Motivation, Motivation .....

We now embark on a study of non-relativistic quantum mechanics.

Quantum mechanics explains all physical systems when the characteristic length scale of the phys- ical system is smaller than 10^{-4} m.

We will not endeavor to explain why nature works in this particular way since it is my strong belief, as we will see in these lectures, that it is not possible to do so within the context of theoretical physics.

Let me emphasize my strong belief that theories in physics should only endeavor to make predictions about experimental measurements and not attempt to provide reasons for why nature works in these particular ways, that is, why we must choose to start with certain postulates.

We will be able to correctly predict the outcomes of an amazingly wide range of experiments.

In order to understand the ideas and implications of the theory we will have to adjust all of our ways of thinking at the most fundamental level - leave you classical intuition behind!

The language of quantum mechanics is mathematics.

We will develop the mathematics in the so-called Dirac language, which, as we will see, is a completely new way of thinking about quantum mechanics.

We covered most of the necessary mathematics in the Mathematical Methods course offered last winter.

The major missing topic is linear algebra.

There are note on the website to bring you up to speed on those aspects of linear algebra that we will need.

Mostly, however, you will learn linear algebra as we use it in lecture and do homework problems.

Chapter 2

The Mathematics of Quantum Physics: Dirac Language

I will assume that you have already read the chapter before I lecture and have, therefore, learned many of the ideas.

If I do not cover a topic in the book in lecture but you are struggling to understand it, just ask a question in lecture.

Mappings

Given two sets A and B, let a $\in$ A, b $\in$ B.

A mapping T from A to B is defined by:

$$a \mapsto b = Ta$$

Details of type are in the book.

Linear Vector Spaces

A vector space is defined with reference to a field.

We consider the case where the field is the field C of complex numbers because this is the case of interest in quantum mechanics.

Ket Vectors

The mathematical vectors that will allow us to describe physical states will be called **ket vectors** or **kets** and be represented by the symbol $|...\rangle$ (due to Dirac).

We will label different vectors(states) according to their associated physical(measurable) properties and these will be inserted inside the ket symbol $|a,b,...\rangle$.

The ket vectors will form the basis for the Dirac language of the associated vector space defined below.

At this point let us state a couple of the basic properties of the ket vectors and their connection to physics to help set the stage for the discussions that will follow.

Kets can be multiplied by complex numbers and can be added together to get another ket vector

$$c_1 |a\rangle + c_2 |b\rangle = |q\rangle$$

complex numbers

The crucial assumptions we will make later as we connect the Dirac language, the mathematical formalism and the physics are:

- Each state of a physical system at a given time can be mathematically described by a definite ket vector. There is some correspondence!
- If a physical state is a superposition(sum) of other physical states, its corresponding ket vector is a linear sum(combination) of the ket vectors corresponding to the other states

The state |q⟩ in (2.2) is a **superposition** of the states |a⟩ and |b⟩ with the mathematical properties of this superposition defined precisely by the two complex numbers c_1 and c_2.

For example, the state of a photon passing through a double-slit apparatus might be described by the superposition

$$|photon\ in\ 2-slit\ apparatus\rangle = a\,|slit\ 1\rangle + b\,|slit\ 2\rangle$$

where |slit 1⟩ is the state of a photon that passes through slit 1.

Now for the mathematical formalism.

A **linear vector space** *V* is a set of abstract elements, called vectors, for which there exists two rules:

1. a rule for creating a vector sum or vector addition,

$$|7\rangle + |12\rangle$$

2. a rule for multiplication of a vector by a complex scalar c,

$$c\,|13\rangle$$

The following properties hold:

1. the result of applying either or both of these rules is another vector in the same space; this is called *closure*

$$c_1\,|7\rangle + c_2\,|12\rangle \in V$$

2. scalar multiplication is *distributive* with respect to the vectors

$$c(|7\rangle + |12\rangle) = c|7\rangle + c|12\rangle$$

3. scalar multiplication is *associative*

$$c_1(c_2|7\rangle) = c_1 c_2|7\rangle$$

4. addition is *commutative*

$$|7\rangle + |12\rangle = |12\rangle + |7\rangle$$

5. addition is *associative*

$$|6\rangle + (|7\rangle + |12\rangle) = (|6\rangle + |7\rangle) + |12\rangle$$

6. there exists a unique *zero* or *null* vector $|\varnothing\rangle$ such that

$$|7\rangle + |\varnothing\rangle = |7\rangle \text{ and } 0|M\rangle = |\varnothing\rangle$$

7. for every vector $|M\rangle$ there exists a unique additive *inverse* vector $|-M\rangle$ where

$$|M\rangle + |-M\rangle = |\varnothing\rangle \text{ and } |-M\rangle = -|M\rangle$$

We must emphasize that these "vectors" are abstract mathematical objects whose properties are defined above - forget the old”magnitude and direction idea!

Example: Consider the set of all 3-tuples, which is a particular example of a finite-dimensional vector space

$$|i\rangle = \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} \qquad |i\rangle + |i\rangle = \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} + \begin{pmatrix} a_j \\ b_j \\ c_j \end{pmatrix} = \begin{pmatrix} a_i + a_j \\ b_i + b_j \\ c_i + c_j \end{pmatrix} \qquad q\,|i\rangle = \begin{pmatrix} qa_i \\ qb_i \\ qc_i \end{pmatrix} \qquad |\varnothing\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Example: the set of all functions *f(x)* of a real variable *x* for which

$$\int |f(x)|^2 \, dx < \infty$$

with addition and multiplication by a scalar defined by

$$(f+g)(x) = f(x) + g(x) \; and \; (af)(x) = af(x)$$

This space is called L^2.

Example: the set of all infinite sequences of numbers $x = (x_1, x_2, x_3, \ldots\ldots, x_i, \ldots..)$ such that

$$x + y = (x_1 + y_1, x_2 + y_2, x_3 + y_3, \ldots, x_k + y_k, \ldots)$$

with addition and multiplication by a scalar defined by

$$(f+g)(x) = f(x) + g(x) \qquad (af)(x) = af(x)$$

$$ax = (ax_1, ax_2, ax_3, \ldots, ax_k, \ldots)$$

Example: the set of all 2 x 2 matrices $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$

with addition and multiplication by a scalar defined by

$$A + B = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$$

$$cA = c\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{pmatrix}$$

Definition: A set of vectors is said to be **linearly independent** if a linear relation of the form

$$\sum_{k=1}^{n} c_k |k\rangle = |\varnothing\rangle \quad = 0 \text{ for simplicity.}$$

implies that $c_k = 0$ for all k; otherwise the set of vectors is linearly dependent.

If a set of vectors is linearly dependent, then we can express a member of the set as a linear combination of the other members of the set.

Example: Consider the set of vectors(3-tuples in this case)

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

This set is linearly independent since

$$a_1 |1\rangle + a_2 |2\rangle + a_3 |3\rangle = \begin{pmatrix} a_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = |\varnothing\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that the only solution is

$$a_1 = a_2 = a_3 = 0$$

Example: Consider the set of vectors(3-tuples in this case)

$$|1\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

This set is linearly independent since

$$a_1 |1\rangle + a_2 |2\rangle + a_3 |3\rangle = \begin{pmatrix} a_1 \\ a_1 \\ 0 \end{pmatrix} + \begin{pmatrix} a_2 \\ -a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 \\ a_1 - a_2 \\ a_3 \end{pmatrix} = |\varnothing\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that the only solution is

$$a_1 + a_2 = a_1 - a_2 = a_3 = 0 \qquad \text{or} \qquad a_1 = a_2 = a_3 = 0$$

Example: Consider the set of vectors(3-tuples in this case)

$$|1\rangle = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

This set is linearly independent since

$$a_1 |1\rangle + a_2 |2\rangle + a_3 |3\rangle = \begin{pmatrix} a_1 \\ a_1 \\ a_1 \end{pmatrix} + \begin{pmatrix} a_2 \\ -a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 2a_3 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 + 2a_3 \\ a_1 - a_2 \\ a_1 + a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that the solution is

$$a_1 + a_2 + 2a_3 = a_1 - a_2 = a_1 + a_3 = 0 \qquad \text{or} \qquad a_2 = a_1, a_3 = -a_1$$

Definition: The maximum number of linearly independent vectors in a space V is called the **dimension** of the space $dim(V)$.

Definition: A set of vectors $|k\rangle$, $k = 1,2,3,...,n$ spans the space if every vector $|Q\rangle$ in the space can be written as a linear combination of vectors in the set

$$|Q\rangle = \sum_{k=1}^{n} q_k \, |k\rangle$$

This linear combination, which is given by the coefficients q_k,$k = 1,2,...,n$ is unique.

Definition: A set of vectors is a basis for the space if it is a linearly independent set and spans the space, that is, if dim(V) = m, a set of m linearly independent vectors is called a *basis* on V.

The set of vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

is the **maximal** set of linearly independent vectors since any other vector $|g\rangle$ in the space can always be written as a linear combination of them as

$$|g\rangle = a_1 \, |1\rangle + a_2 \, |2\rangle + a_3 \, |3\rangle = \begin{pmatrix} a_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

Therefore the dimension of this vector space is 3.

This set of vectors is also a basis.

The basis is **not unique** since the set of linearly independent vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

also spans the space, i.e.,

$$|g\rangle = c_1 |1\rangle + c_2 |2\rangle + c_3 |3\rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} c_1 + c_2 \\ c_1 + c_2 \\ c_3 \end{pmatrix}$$

implies that

$$2c_1 = a_1 + a_2 \quad 2c_2 = a_1 - a_2 \quad c_3 = a_3$$

and, thus, this set is also a basis.

Clearly, a basis spans the whole of V.

Definition: The coefficients in the expansion of an arbitrary vector $|Q\rangle$ in terms of a basis set $|k\rangle$, $k = 1,2,3,...,n$

$$|Q\rangle = \sum_{k=1}^{n} q_k |k\rangle$$

are called the **components** of the vector $|Q\rangle$ in that basis.

Example: In the space of 3-tuples, a basis is represented by the three vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

so that an arbitrary vector in the space can be written

$$|g\rangle = a_1|1\rangle + a_2|2\rangle + a_3|3\rangle = \begin{pmatrix} a_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

so that a_1, a_2 and a_3 are the components.

When we add vectors(must be defined with respect to the same basis) we simply add their components

$$|Q\rangle + |R\rangle = \sum_{k=1}^{n} q_k|k\rangle + \sum_{k=1}^{n} r_k|k\rangle = \sum_{k=1}^{n} (q_k + r_k)|k\rangle$$

Inner Products

We can, however, define generalizations of length and direction in an arbitrary space using a mathematical object called the inner product.

The inner product is a generalization of the standard dot product.

I will first discuss the inner product using standard mathematical notation and then return to the Dirac language.

Definition: An **inner product** for a linear vector space associates a scalar (f, g) with every **ordered pair** of vectors f, g.

It satisfies these properties:

1.

$$(f, g) = \text{ complex number}$$

2.

$$(f, g) = (g, f)^*$$

3.

$$(f, c_1 g_1 + c_2 g_2) = c_1 (f, g_1) + c_2 (f, g_2)$$

4.

$$(f, f) \geq 0 \text{ with equality if and only if } f = \text{ null vector}$$

Now(2) and (3) above imply that

$$(c_1 g_1 + c_2 g_2, f) = c_1^* (f, g_1) + c_2^* (f, g_2)$$

Hence, the inner product is said to be **linear** in its **second** argument and **anti-linear** in its **first** argument.

Definition: The non-negative number $\|f\| = (f, f)^{1/2}$ is called the *norm* or *length* of the vector f. Clearly, $\|f\| = 0$ if and only if $f = 0$.

Definition: If the inner product of two vectors is zero, then the vectors are *orthogonal.*

Definition: A set of vectors $\{f_i\}$ is said to be *orthonormal* if the vectors are pairwise orthogonal and of unit norm. We can summarize this property by the equation

$$(f_i, f_j) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

where the symbol δ_{ij} is called the *Kronecker delta.*

Schwarz's Inequality: scalar products satisfy

$$|(f,g)|^2 \leq (f,f)(g,g)$$

Triangle Inequality: scalar products satisfy

$$\|(f+g)\| \leq \|f\| + \|g\|$$

Equality holds in both cases only if one vector is a scalar multiple of the other, i.e., $f = cg$ where the scalar c is real and positive.

An inner product space is simply one in which an inner product is defined.

Examples:

1. For the case of n-tuples(an n-dimensional vector space)

$$f = (x_1, x_2, x_3, \ldots, x_n), \quad , \quad g = (y_1, y_2, y_3, \ldots, y_n)$$

$$(f,g) = \sum_{k=1}^{n} x_k^* y_k$$

2. For the case of square integrable functions

$$(f,g) = \int f^*(x) g(x) w(x)\, dx$$

where $w(x)$ is a nonnegative weight function.

When a vector Q is expanded as a linear combination

$$Q = \sum_k q_k f_k$$

of orthonormal basis vectors f_k, the coefficients or components are given by the inner product, i.e.,

$$(f_j, Q) = \sum_k q_k (f_j, f_k) = \sum_k \delta_{kj} = q_j$$

or

$$Q = \sum_k (f_k, Q) f_k$$

Of the two basis vector sets for the 3-dimensional vectors we looked at earlier:

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

is an orthonormal basis and

$$|1\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

is orthogonal but not orthonormal.

Given two inner product spaces U and V, their direct sum $W = U \oplus V$ is the inner product space consisting of the pairs $\{f.g\}$ written $f \oplus g, f \in U, g \in V$ with the operations defined by

$$\alpha_1 f_1 \oplus g_1 + \alpha_2 f_2 \oplus g2 = (\alpha_1 f_1 + \alpha_2 f_2) \oplus (\alpha_1 g_1 + \alpha_2 g_2)$$

$$(f_1 \oplus g_1, f_2 \oplus g2) = (f_1, f_2) + (g_1, g_2)$$

Linear Functionals

Linear complex-valued functions of vectors are called **linear functionals** on V.

A linear functional Γ assigns a scalar value $\Gamma(f)$ to each vector f in the vector space V, such that linearity holds

$$\Gamma(c_1 f_1 + c_2 f_2) = c_1 \Gamma(f_1) + c_2 \Gamma(f_2)$$

for any vectors f_1 and f_2 and any scalars c_1 and c_2.

The set of linear functionals on V form a linear space V' where we define addition and scalar multiplication by

$$(\Gamma_1 + \Gamma_2)(f) = \Gamma_1(f) + \Gamma_2(f)$$

$$(c\Gamma)(f) = c\Gamma(f)$$

The vector space V' is called the **dual space** of the vector space V.

There is a one-to-one correspondence between linear functionals Γ in V' and the vectors f in V, such that the linear functional associated with f has the form

$$\Gamma_f(g) = (f, g)$$

where f is a fixed vector and g is an arbitrary vector.

This is called the Riesz theorem.

Using the properties of the inner product we can show that

$$\Gamma_f + \Gamma_g = \Gamma_{f+g} \qquad \text{and} \qquad a\Gamma_f = \Gamma_{a^* f}$$

or

$$\Gamma_f(h) + \Gamma_g(h) = \Gamma_{(f+g)}(h)$$

$$(f, h) + (g, h) = (f + g, h)$$

and

$$a\Gamma_f(h) = a(f,h) = (a^* f, h) = \Gamma_{a^* f}$$

The scalar product is clearly **antilinear** in Γ and linear in f.

The vector f that corresponds to a given linear functional Γ_f is easily found by direct construction.

$$f = \sum_n [\Gamma_f(\alpha_n)]^* \alpha_n$$

See book for derivation.

The Dirac language of Kets and Bras

We now rewrite everything we have done so far in the Dirac language.

Vectors in the linear space V are called **kets** or **ket vectors** and denoted by $|a\rangle$.

Linear functionals Γ_b in the dual linear space V' are called **bras** or **bra vectors** and are denoted by $\langle b|$.

There is a one-to-one correspondence between kets in V and bras in V' or between vectors and linear functionals.

We will use the same label to denote these corresponding quantities

$$|a\rangle \leftrightarrow \langle a|$$

The inner product between two vectors $|a\rangle$ and $|b\rangle$ in V corresponds to the linear functional of the left-hand vector assigning a scalar value to the right-hand vector

$$\Gamma_a(b) = (a,b) = \langle a \,|\, b\rangle = \text{ complex number}$$

The orthonormality property of a basis set $\{\alpha_i\}$ is expressed by the linear functional relation

$$\langle \alpha_i \,|\, \alpha_j \rangle = \delta_{ij}$$

The expansion of an arbitrary state in an orthonormal basis is given by

$$|q\rangle = \sum_n a_n |\alpha_n\rangle$$

and the expansion coefficients are

$$\langle \alpha_m | q \rangle = \sum_n a_n \langle \alpha_m | \alpha_n \rangle = \sum_n a_n \delta_{mn} = a_m$$

or a_m is the linear functional $\Gamma_{\alpha m}(q)$ of the state $|q\rangle$ with respect to the corresponding basis vector $|\alpha_m\rangle$.

If we have two vectors $|a\rangle$ and $|b\rangle$ and we expand them in an orthonormal basis set $\{\alpha_i\}$ as

$$|a\rangle = \sum_n a_n |\alpha_n\rangle \text{ and } |b\rangle = \sum_n b_n |\alpha_n\rangle$$

$$\rightarrow \langle a | b \rangle = \sum_n a_n^* b_n = \langle b | a \rangle^* = \left(\sum_n b_n^* a_n \right)^*$$

where we have used the antilinear property of inner products that says that we have the correspondence

$$c|a\rangle \leftrightarrow c^* \langle a|$$

The linear functional itself is directly represented by the bra vector, i.e.,

$$\langle q| \cdots = \Gamma_q(\cdots) = (q, \cdots)$$

In Dirac language, there is nothing to distinguish between the value of the linear functional $\langle q|$ for the vector $|p\rangle$ and the inner product of the vectors $|q\rangle$ and $|p\rangle$.

They are both $\langle q \mid p \rangle$.

Gram-Schmidt Orthogonalization Process

Full derivation in book.

An orthonormal basis set for an n-dimensional vector space can always be constructed from any set of n linearly independent vectors using the Gram-Schmidt orthogonalization method.

Suppose that we have a set of n linearly independent vectors $|\alpha_i\rangle$, $i = 1, 2, \ldots, n$ that are not a mutually orthonormal set. We can construct a mutually orthonormal set $|\beta_i\rangle$, $i = 1, 2,\ldots, n$ using the following rules:

$$|\beta_1\rangle = |\alpha_1\rangle$$

$$|\beta_{k+1}\rangle = |\alpha_{k+1}\rangle + \sum_{j=1}^{k} a_j |\beta_j\rangle \qquad a_j = -\frac{\langle \beta_j | \alpha_{k+1}\rangle}{\langle \beta_j | \beta j\rangle}$$

Example in book.

Linear Operators

In general, an operator defined on a vector space maps vectors into vectors $\hat{Q} : V \rightarrow V$, i.e., if $\hat{Q}$ = operator and $|k\rangle$= a vector, then $\hat{Q}\ |\mathrm{k}\rangle = |\mathrm{q}\rangle$= a vector (we will use the ^ symbol on top to signify operators).

We say that $|q\rangle$ is the image of $|k\rangle$ under the mapping $\hat{Q}$.

An operator is fully defined if we specify its action on all vectors in its domain, which is the set of all vectors for which $\hat{Q}\ |k\rangle$ for all $|k\rangle$ is defined.

Its range is the set of all vectors of the form $\hat{Q}\ |k\rangle$ for all $|k\rangle$.

In the special case of a **linear operator**, the operator also satisfies the **linearity relation**

$$\hat{Q}(a_1 |k_1\rangle + a_2 |k_2\rangle) = \hat{Q}(a_1 |k_1\rangle) + \hat{Q}(a_2 |k_2\rangle)$$

In general, the operator $\hat{A}$ is linear if, for all $|f\rangle, |g\rangle \in V$, $\hat{A}(|f\rangle + |g\rangle) = \hat{A}|f\rangle + \hat{A}|g\rangle$, $(\alpha \hat{A})|f\rangle = (\alpha \hat{A}|f\rangle$ and $\hat{A} = 0$ if and only if $\hat{A}|f\rangle = 0$ for all $|f\rangle$.

The norm $\|\hat{A}\|$ of the linear operator $\hat{A}$ is defined as the maximum of $\|\hat{A}|f\rangle\|$ for all $|f\rangle$ such that $\||f\rangle\| \leq 1$.

The identity operator $\hat{I}$ is defined by $\hat{I}|f\rangle = |f\rangle$ for all $|f\rangle$.

Since any vector in the space can be written as a linear combination of a basis set of vectors, we need only define a linear operator on the basis set and then its operation on all vectors is known.

Quantum mechanics works exclusively with linear operators so we will just call them operators from now on.

We define the sum and product of operators by

$$(\hat{Q}_1 + \hat{Q}_2)|k\rangle = \hat{Q}_1 |k\rangle + \hat{Q}_2 |k\rangle$$

$$(\hat{Q}_1 \hat{Q}_2)|k\rangle = \hat{Q}_1(\hat{Q}_2 |k\rangle)$$

for all vectors $|k\rangle$

Associativity is expressed by

$$(\hat{Q}_1 \hat{Q}_2)\hat{Q}_3 = \hat{Q}_1(\hat{Q}_2 \hat{Q}_3)$$

In general we have $\hat{Q}_1 \hat{Q}_2 \neq \hat{Q}_2 \hat{Q}_1$, which corresponds to the **noncommutivity** of the two operators.

We define the commutator of two operators by

$$[\hat{Q}_1, \hat{Q}_2] = \hat{Q}_1 \hat{Q}_2 - \hat{Q}_2 \hat{Q}_1$$

Two operators are said to commute if their corresponding commutator is zero (the null vector).

The commutator will be a fundamental quantity in our discussions of quantum mechanics.

An Aside : The Connection to Matrices

Now suppose that we have an *N*-dimensional vector space and we choose an orthonormal basis set

$$\{|q_i\rangle, i = 1, 2, \ldots, N\}$$

In an *N*-dimensional space, we can always represent any vector as an *N*-element column of numbers or as a column vector.

If we expand some arbitrary state $|\alpha\rangle$ in terms of this basis set and operate on it with an operator $\hat{Q}$, then we have

$$|\beta\rangle = \hat{Q}|\alpha\rangle = \hat{Q}\sum_{i=1}^{N} c_i |q_i\rangle = \sum_{i=1}^{N} c_i \hat{Q}|q_i\rangle$$

where $|\beta\rangle$ is another vector in the same space since $\hat{Q}$ is a linear operator.

Now expanding the vector $|\beta\rangle$ in the same basis set as

$$|\beta\rangle = \sum_{i=1}^{N} d_i |q_i\rangle$$

and constructing the linear functional(inner product) of $|\beta\rangle = \hat{Q}|\alpha\rangle$ with respect to $|q_i\rangle$ we get

$$\langle q_k|\hat{Q}|\alpha\rangle = \langle q_k|\hat{Q}\sum_{i=1}^{N} c_i |q_i\rangle = \sum_{i=1}^{N} c_i \langle q_k|\hat{Q}|q_i\rangle = \langle q_k|\beta\rangle = \sum_{j=1}^{N} d_j \langle q_k|q_j\rangle$$

Using the orthonormality property of the basis set

$$\langle q_k | q_j \rangle = \delta_{kj}$$

we get

$$\langle q_k | \hat{Q} | q_i \rangle = \sum_{j=1}^{N} d_j \delta_{kj} = d_k$$

If we define the numbers $\hat{Q}_{ki} = \langle q_k | \hat{Q} | q_i \rangle$ we get the relation

$$\sum_{j=1}^{N} \hat{Q}_{ki} c_i = d_k$$

Clearly, this has the form of a matrix equation where $\hat{Q}_{ki} = \langle q_k | \hat{Q} | q_i \rangle$ is defined as the (*ki*)-matrix element of the operator $\hat{Q}$.

In this case, we would have the representation

$$(\hat{Q}_{ij}) = N \times N \text{ matrix}$$

$$(c_i) = N \times 1 \text{ column matrix}$$

$$(d_i) = N \times 1 \text{ column matrix}$$

and the matrix equation (using matrix multiplication) represented above is $Qc = d$.

All finite dimensional vector space operator equations can be turned into matrix equations in this manner.

More about Vectors, Linear Functionals, Operators

We discussed what happens when an operator acts to the right on a ket vector.

What about acting to the left on a bra vector, i.e., what is the meaning of the quantity

$$\langle q | \hat{Q}$$

The bra vector is really a linear functional

Remember that the definition of the linear functional was

$$\langle q \,|\, p \rangle = \Gamma_q(p) = (q, p)$$

The standard definition (in terms of inner products) of the adjoint operator $\hat{Q}^\dagger$, of the operator $\hat{Q}$, in a linear vector space is

$$(\hat{Q}^\dagger q, p) = (q, \hat{Q} p)$$

The adjoint obeys the following properties:

$$\|\hat{A}^\dagger\| = \|\hat{A}\| \qquad (\hat{A}^\dagger)^\dagger = \hat{A} \qquad (\hat{A} + \hat{B})^\dagger = \hat{A}^\dagger + \hat{B}^\dagger$$

$$(\alpha \hat{A})^\dagger = \alpha^* \hat{A}^\dagger \qquad (\hat{A}\hat{B})^\dagger = \hat{B}^\dagger \hat{A}^\dagger \qquad (\hat{A}^{-1})^\dagger = (\hat{A}^\dagger)^{-1}$$

If we define a new vector ϕ by the operation $\hat{Q}^\dagger q = \phi$, then using the definition of the adjoint operator we have

$$\Gamma_\phi(p) = (\phi, p) = (\hat{Q}^\dagger q, p) = (q, \hat{Q} p) = \Gamma_q(\hat{Q} p)$$

We want the action of an operator on the bra space of linear functionals or bra vectors to create a new linear functional or bra vector in the same way that it does for ket vectors.

We can guarantee this by **defining** the operation of an operator on the bra space of linear functionals by

$$\hat{Q}\Gamma_q(p) = \Gamma_q(\hat{Q} p)$$

Using this new definition and the definition of a linear functional, we then have

$$\hat{Q}\Gamma_q(c_1p_1 + c_2p_2) = \Gamma_q(\hat{Q}(c_1p_1 + c_2p_2))$$

$$= c_1\Gamma_q(\hat{Q}p_1) + c_2\Gamma_q(\hat{Q}p_2)$$

$$= c_1\hat{Q}\Gamma_q(p_1) + c_2\hat{Q}\Gamma_q(p_2)$$

which says that $\hat{Q}\ \Gamma_q(...)$ itself is also a linear functional(another bra vector).

Thus, our definition of how an operator acts on the bra space of linear functionals simply says that it creates a new functional as we desired.

With this definition, we then have

$$\hat{Q}\Gamma_q(p) = \Gamma_\phi(p) = (\phi, p) = (\hat{Q}^\dagger q, p) = (q, \hat{Q}p) = \Gamma_q(\hat{Q}p)$$

or, since $(\hat{Q}^\dagger q, p) = \Gamma_{\hat{Q}^\dagger q}(p)$, the relationship among linear functionals is

$$\hat{Q}\Gamma_q(\ldots) = \Gamma_{\hat{Q}^\dagger q}(p)$$

In terms of bra and ket vectors, this says that if $\langle q|$ and $|q\rangle$ are corresponding bra and ket vectors, then

$$\hat{Q}\,|q\rangle = |\beta\rangle \quad \text{and} \quad \langle q|\,\hat{Q}^\dagger = \langle\beta|$$

Since $\langle\beta|p\rangle^* = \langle p|\beta\rangle$ we then have that

$$\langle q|\,\hat{Q}^\dagger\,|p\rangle^* = \langle p|\,\hat{Q}\,|q\rangle$$

This relation is equivalent to the original inner product relation

$$(\hat{Q}^\dagger q, p) = (q, \hat{Q}p)$$

End result of discussion is that $\langle q| \hat{Q}^\dagger = \langle \beta|$ is the bra vector(linear functional) corresponding to the ket vector $\hat{Q}\ |q\rangle = |\beta\rangle$.

Since the adjoint operator satisfies the relations

$$(c\hat{Q})^\dagger = c^*\hat{Q}^\dagger \qquad (\hat{Q}\hat{R})^\dagger = \hat{R}^\dagger\hat{Q}^\dagger \qquad (\hat{Q}+\hat{R})^\dagger = \hat{Q}^\dagger + \hat{R}^\dagger$$

we can define another product among the bra and ket vectors, which is an operator rather than a scalar as in the case of the inner product.

It is called the **outer product** and, in the Dirac language, has the form

$$|q\rangle\langle p|$$

Outer product is an operator since its action on any other vector always results in a vector

$$(|q\rangle\langle p|)\,|s\rangle = |q\rangle\langle p\,|\,s\rangle$$

We also have that

$$(|q\rangle\langle p|)^\dagger = |q\rangle\langle p|$$

see book for derivation

Operator is very important in our later discussions.

The special case

$$|p\rangle\langle p|$$

is called a **projection operator** because it picks out the component of an arbitrary state vector in the "direction" of the vector $|p\rangle$.

Projections operators will be the mathematical object with a direct connection to physical states and measurements in our later discussions.

Example in a Finite Dimensional Vector Space

Consider the 2-dimensional vector space spanned by the orthonormal basis set

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Define two projection operators

$$\hat{P}_1 = |1\rangle\langle 1| \qquad \hat{P}_2 = |2\rangle\langle 2|$$

We have

$$(\hat{P}_1) = \begin{pmatrix} \langle 1|\hat{P}_1|1\rangle & \langle 1|\hat{P}_1|2\rangle \\ \langle 2|\hat{P}_1|1\rangle & \langle 2|\hat{P}_1|2\rangle \end{pmatrix}$$

$$= \begin{pmatrix} \langle 1|1\rangle\langle 1|1\rangle & \langle 1|1\rangle\langle 1|2\rangle \\ \langle 2|1\rangle\langle 1|1\rangle & \langle 2|1\rangle\langle 1|2\rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

matrix representation

$$(\hat{P}_2) = \begin{pmatrix} \langle 1|\hat{P}_2|1\rangle & \langle 1|\hat{P}_2|2\rangle \\ \langle 2|\hat{P}_2|1\rangle & \langle 2|\hat{P}_2|2\rangle \end{pmatrix}$$

$$= \begin{pmatrix} \langle 1|2\rangle\langle 2|1\rangle & \langle 1|2\rangle\langle 2|2\rangle \\ \langle 2|2\rangle\langle 2|1\rangle & \langle 2|2\rangle\langle 2|2\rangle \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Arbitrary vector in this space

$$|a\rangle = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1|1\rangle + a_2|2\rangle$$

We then have (using both Dirac and matrix languages)

$$\hat{P}_1 |a\rangle = a_1 |1\rangle \langle 1|1\rangle + a_2 |1\rangle \langle 1|2\rangle = a_1 |1\rangle$$

$$(\hat{P}_1)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ 0 \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

as expected

Note that

$$(\hat{P}_1) + (\hat{P}_2) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \hat{I} = \text{identity operator}$$

$$(\hat{P}_1) + (\hat{P}_2) |a\rangle = (\langle 1|1\rangle + \langle 2|2\rangle) |a\rangle$$

$$= \sum_{j=1}^{2} |j\rangle \langle j|a\rangle = |a\rangle = \hat{I} |a\rangle$$

Return to Gram-Schmidt

As before, suppose we have a set of linearly independent, but non-orthogonal vectors $|i\rangle$ in an n-dimensional linear vector space, we can construct a set of orthogonal vectors $|\alpha_i\rangle$ as follows:

$$|\alpha_1\rangle = |1\rangle \qquad |\alpha_2\rangle = |2\rangle - |\alpha_1\rangle \langle \alpha_1|2\rangle$$

$|\alpha_1\rangle \langle \alpha_1| = \hat{P}_{\alpha_1}$ = projection operator $\longrightarrow$ $|\alpha_2\rangle = |2\rangle - \hat{P}_{\alpha_1} |2\rangle$

Now consider

$$|Q\rangle = q_1 |1\rangle + q_2 |2\rangle + q_3 |3\rangle + \ldots = \sum_i q_i |i\rangle$$

$$\hat{P}_{\alpha_n} = |\alpha_n\rangle \langle\alpha_n|$$

$$\hat{P}_{\alpha_n} |Q\rangle = \sum_i q_i \hat{P}_{\alpha_n} |i\rangle = \sum_i q_i |\alpha_n\rangle \langle\alpha_n | i\rangle = \sum_i q_i |\alpha_n\rangle \delta_{ni} = q_n |\alpha_n\rangle$$

$$\sum_n \hat{P}_{\alpha_n} |Q\rangle = \sum_n q_n |\alpha_n\rangle = |Q\rangle$$

$$\longrightarrow \qquad \sum_n |\alpha_n\rangle \langle\alpha_n| = \hat{I} = \text{ identity operator}$$

which makes perfect sense since if you project out all components of a vector you just get the vector back!

This generalizes the earlier result we found in the 2-dimensional space.

This Dirac language is very interesting!

We can then write a general Gram-Schmidt procedure as

$$|\alpha_1\rangle = |1\rangle \qquad |\alpha_2\rangle = (1 - \hat{P}_{\alpha_1}) |2\rangle \qquad |\alpha_3\rangle = (1 - \hat{P}_{\alpha_1} - \hat{P}_{\alpha_2}) |3\rangle$$

.....

$$|\alpha_n\rangle = \left(1 - \sum_{k=1}^{n-1} \hat{P}_{\alpha_k}\right) |n\rangle$$

which is rather neat.

This shows the power of these projection operators and of the Dirac language.

Finally, we define a property of an operator called the **trace** as

$$Tr\hat{Q} = \sum_j \langle q_j | \hat{Q} | q_j \rangle = \text{ sum of diagonal elements } = \sum_j (\hat{Q})_{jj}$$

Self-Adjoint Operators (details in book)

If

$$\langle q | \hat{Q}^\dagger | p \rangle = (\langle p | \hat{Q} | q \rangle)^*$$

then $\hat{Q}$ is a **Hermitian or self-adjoint operator**.

Hermitian operator matrix elements in a finite-dimensional space are

$$\hat{Q}^*_{ki} = \hat{Q}_{ik} \qquad or \qquad \hat{Q}^\dagger = \hat{Q}$$

For the Hermitian operator $\hat{H}$

$$|(f, \hat{H} g)|^2 \geq (f, \hat{H} f)(g, \hat{H} g)$$ Schwarz inequality

If $\hat{K}^\dagger = -\hat{K}$, then the operator $\hat{K}$ is called **antihermitian**; clearly, i $\hat{K}$ is then Hermitian.

An arbitrary operator $\hat{A}$ can always be written as the sum $\hat{A} = \hat{H}_A + \hat{K}_A$ of its Hermitian part

$$\hat{H}_A = \frac{1}{2}(\hat{A} + \hat{A}^\dagger)$$

and its antihermitian part

$$\hat{K}_A = \frac{1}{2}(\hat{A} - \hat{A}^\dagger)$$

Hermitian operators will be very important in our discussions because they will represent measurable quantities or observables.

Eigenvalues and Eigenvectors (details in book)

Suppose

$$\hat{Q}\,|\beta\rangle = b\,|\beta\rangle$$

The vector $|\beta\rangle$ is then called an eigenvector and the scalar b an eigenvalue of the operator $\hat{Q}$.

We then also must have (using any arbitrary vector $|\alpha\rangle$)

$$\langle\alpha|\,\hat{Q}\,|\beta\rangle^* = b^*\,\langle\alpha\,|\,\beta\rangle^*$$

$$\langle\beta|\,\hat{Q}^\dagger\,|\alpha\rangle = b^*\,\langle\beta\,|\,\alpha\rangle$$

$$\langle\beta|\,\hat{Q}^\dagger = b^*\,\langle\beta|$$

Assume that $\hat{Q}$ is a Hermitian operator, then

$$\langle\beta|\,\hat{Q}^\dagger\,|\beta\rangle = \langle\beta|\,\hat{Q}\,|\beta\rangle = \langle\beta|\,\hat{Q}\,|\beta\rangle^*$$

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

$$(b - b^*)\,\langle\beta\,|\,\beta\rangle = 0$$

$$b = b^*$$

This means that all of the eigenvalues of a Hermitian operator are **real**.

In addition,

$$\langle\beta|\,\hat{Q} = b\,\langle\beta|$$

or the ket vector and its corresponding bra vector are eigenvectors with the same eigenvalue.

Suppose that we have two eigenvectors

$$\hat{Q}\,|\alpha\rangle = a\,|\alpha\rangle \quad and \quad \hat{Q}\,|\beta\rangle = b\,|\beta\rangle$$

then

$$0 = \langle\alpha|\,\hat{Q}\,|\beta\rangle - \langle\beta|\,\hat{Q}^{\dagger}\,|\alpha\rangle^{*} = \langle\alpha|\,\hat{Q}\,|\beta\rangle - \langle\beta|\,\hat{Q}\,|\alpha\rangle^{*}$$

$$0 = b\,\langle\alpha\,|\,\beta\rangle - a^{*}\,\langle\beta\,|\,\alpha\rangle^{*} = (b-a)\,\langle\alpha\,|\,\beta\rangle$$

This implies that eigenvectors corresponding to distinct(different) eigenvalues ($a \neq b$) are orthogonal, i.e., $\langle\alpha \mid \beta\rangle = 0$.

If $a = b$, that is, the two eigenvectors correspond to the same eigenvalue, then they are called **degenerate** eigenvectors.

We will always assume that we have already carried out this orthogonalization process and that the set of eigenvectors(for both non-degenerate and degenerate eigenvalues) of any Hermitian operator is an orthonormal set where.

$$\langle\alpha_i\,|\,\alpha_j\rangle = \delta_{ij}$$

The set of eigenvalues of an operator $\hat{A}$ is called the **spectrum** of $\hat{A}$.

In an m-dimensional space $\mathcal{H}$ choose a matrix representation in which operator $\hat{A}$ = (A_{ij}) and the vector g = (γ_i).

Then the eigenvalue equation $\hat{A}g = \lambda g$ becomes a system of m homogeneous linear equations in the γ_i.

The eigenvalues λ are the roots of the algebraic equation of degree m

$$det(A_{ij} - \lambda\delta_{ij}) = 0$$

known as the **secular equation**, which expresses the condition for the homogeneous system of equations to have non-trivial solutions.

In an *m*-dimensional space every operator has at least one and at most *m* distinct eigenvalues.

If all the eigenvalues of an arbitrary operator $\hat{A}$ are simple, then there are *m* linearly independent eigenvectors of $\hat{A}$ but there may be fewer if $\hat{A}$ has multiple eigenvalues.

Completeness

A set of orthonormal vectors $\{|\alpha_k\rangle, k = 1,2,3,...,N\}$ is **complete** if we can expand an arbitrary vector $|\eta\rangle$ in terms of that set, i.e.,

$$|\eta\rangle = \sum_j a_j |\alpha_j\rangle$$

Then, it can be shown that (see book) in general

$$\sum_j |\alpha_j\rangle \langle\alpha_j| = \hat{I} = \text{ identity operator}$$

for any complete set of orthonormal vectors.

This is one of the most important results we will derive.

It will enable us to perform very clever tricks during algebraic manipulations.

It is fundamentally linked to the probability interpretation of quantum mechanics as we shall see later.

If a set of vectors $\{|q_k\rangle, k = 1,2,3,...,N\}$ are eigenvectors of a Hermitian operator $\hat{Q}$, then we will always assume that they are a complete set.

Thus, if

$$\hat{Q}\,|q_k\rangle = q_k\,|q_k\rangle$$

then we can always write

$$\hat{Q} = \hat{Q}\hat{I} = \hat{Q}\sum_j |q_j\rangle\langle q_j| = \sum_j q_j\,|q_j\rangle\langle q_j|$$

which completely expresses an operator in terms of its eigenvectors and eigenvalues.

In addition, we have

$$\hat{Q}^n = \hat{Q}^n\hat{I} = \hat{Q}^n\sum_j |q_j\rangle\langle q_j| = \sum_j q_j^n\,|q_j\rangle\langle q_j|$$

$$f(x) = \sum_k c_k x^k$$

$$f(\hat{Q}) = f(\hat{Q})\hat{I} = f(\hat{Q})\sum_j |q_j\rangle\langle q_j| = \sum_j\sum_k c_k\hat{Q}^k\,|q_j\rangle\langle q_j|$$

$$= \sum_j\sum_k c_k q_j^k\,|q_j\rangle\langle q_j| = \sum_j f(q_j)\,|q_j\rangle\langle q_j|$$

The remainder of Chapter 2 discusses more mathematical aspects of a numbers of topics; read if you are so inclined - we will use only a few of these items and I will introduce them when needed.

Hilbert Space, Bounded Operators, Inverses, Unitary Operators, More Matrix Representations, Unbounded Operators, Spectra Decomposition-more details, Functions of Operators(Stone's Theorem), Commuting Operators, Continuous Spectrum Operators, Eigenvalues/Eigenvectors of Unitary Operators, Eigenvalue Decomposition, Extension to Infinite-Dimensional Spaces.

Chapter 3

Probability

Probability Concepts

Quantum mechanics involves probability in order to make a connection with experimental measurements.

We will be interested in understanding the quantity

$$P(A|B) = \text{probability of event } A \text{ given that event } B \text{ is true}$$

In essence, event B sets up the conditions or an environment and then we ask about the (conditional) probability of event A given that those conditions exist.

All probabilities are conditional in this sense.

The | symbol means given, so that items to the right of this conditioning symbol are taken as being true.

In other words, we set up an experimental apparatus, which is expressed by properties B and do a measurement with that apparatus, which is expressed by properties A.

We generate numbers (measurements) which we use to give a value to the quantity $P(A|B)$.

Standard Thinking .....

We start with the standard mathematical formalism based on axioms.

We define these events:

1. A = occurrence of A
   (denotes that proposition A is true)
2. $\sim A$ = NOT A = nonoccurrence of A
   (denotes that proposition A is false)
3. $A \cap B$ = A AND B = occurrence of both A and B
   (denotes that proposition A AND B is true)
4. $A \cup B$ = A OR B =occurrence of at least A and B
   (denotes that proposition A OR B is true)

and standard Boolean logic as shown below:

> Boolean logic uses the basic statements AND, OR, and NOT. Using these and a series of Boolean expressions, the final output would be one TRUE or FALSE statement.

This is illustrated below:

1. If A is true AND B is true, then (A AND B) is true
2. If A is true AND B is false, then (A AND B) is false
3. If A is true OR B is false, then (A OR B) is true
4. If A is false OR B is false, then (A OR B) is false

or written as a truth table:

A	B	$(A \cap B)$	$(A \cup B)$
1	1	1	1
1	0	0	1
0	1	0	1
0	0	0	0

Boolean Logic

where 1 = TRUE and 0 = FALSE.

Then we set up a theory of probability with these axioms:

1. $P(A|A) = 1$

   This is the probability of the occurrence A given the occurrence of A. This represents a certainty and, thus, the probability must $= 1$. This is clearly an obvious assumption that we must make if our probability ideas are to make any sense at all.

   In other words, if I set the experimental apparatus such that the meter reads A, then it reads A with probability $= 1$.

2. $0 \leq P(A|B) \leq P(B|B) = 1$

   This just expresses the sensible idea that no probability is greater than the probability of a certainty and it make no sense to have the probability be less than 0.

3. $P(A|B) + P(\sim A|B) = 1$ or $P(\sim A|B) = 1 - P(A|B)$

   This just expresses the fact that the probability of something (anything) happening (A or $\sim A$) given B is a certainty ($= 1$), that is, since the set A or $\sim A$ includes everything that can happen, the total probability that one or the other occurs must be the probability of a certainty and be equal to one.

4. $P(A \cap B|C) = P(A|C)P(B|A \cap C)$

   This says that the probability that 2 events A, B both occur given that C occurs equals the probability of A given C multiplied by the probability of B given $(A \cap C)$, which makes sense which makes sense if you think of them happening in sequence.

All other probability relationships can be derived from these axioms.

The nonoccurrence of A given that A occurs must have probability = 0. This is expressed by

This result clearly follows from the axioms since

$$P(A|B) + P(\sim A|B) = 1$$

$$P(A|A) + P(\sim A|A) = 1$$

$$P(\sim A|A) = 1 - P(A|A) = 1 - 1 = 0$$

In the book, I show an example of using these rules and along the way derive

$$P(A \cup B) = P(A|C) + P(B|C) - P(A \cap B|C)$$

which is a very important and useful result.

If we have $P(A \cap B|C)=0$, then events A and B are said to be mutually exclusive given that C is true and the relation then reduces to

$$P(A \cup B) = P(A|C) + P(B|C)$$

This is the rule of *addition of probabilities for exclusive events*.

Some other important results are:

$$If\ A \cap B = B \cap A, |; then\ P(A|C)P(B|A \cap C) = P(B|C)P(A|B \cap C)$$

$$If\ P(A|C) \neq 0,\ then\ P(B|A \cap C) = P(A|B \cap C)\frac{P(B|C)}{P(A|C)}$$

which is **Baye's theorem**.

It relates the probability of B given A to the probability of A given B.

When we say that B is *independent* of A, we will mean

$$P(B|A \cap C) = P(B|C)$$

or the occurrence of A has NO influence on the probability of B given C.

Using axiom (4) we then have the result:

If A and B are independent given C, then

$$P(A \cap B|C) = P(A|C)P(B|C)$$

This is called **statistical** or **stochastic** independence.

The result generalizes to a set of events $\{A_i, i = 1, 2, \ldots, n\}$.

All these events are independent if and only if

$$P(A_1 \cap A_2 \cap \cdots \cap A_m|C) = P(A_1|C)P(A_2|C)\ldots P(A_m|C)$$

for all $m \leq n$.

Now let us think about these ideas in another way that has fundamental importance in modern approaches to quantum theory. The fundamental result in this view will turn out to be the Bayes formula and its relationship to measurements.

Bayesian Thinking .....

Two Different Axioms

1. If we specify how much we believe something is true, then we must have implicitly specified how much we believe it is false.

2. If we first specify how much we believe that proposition Y is true, and then state how much we believe X is true given that Y is true, then we must implicitly have specified how much we believe that both X and Y are true.

We assign real numbers to each proposition in a manner so that the larger the numerical value associated with a proposition, the more we believe it.

Only using the rules of **Boolean logic**, ordinary algebra, and the constraint that if there are several different ways of using the same information, then we should always arrive at the same conclusions independent of the particular analysis-path chosen, it is then found that this consistency could only be guaranteed if the real numbers we had attached to our beliefs in the various propositions could be **mapped** (or transformed) to another set of real positive numbers which obeyed the usual rules of probability theory:

$$prob(X|I) + prob(\sim X|I) = 1 \quad \text{(same as axiom (3))}$$

$$prob(X \cap Y|I) = prob(X|Y \cap I) \times prob(Y|I) \text{ (same as axiom (4))}$$

The first of these equations is called the sum rule and states (as earlier) that the probability that X is true plus the probability that X is false is equal to one.

The second of these equations is called the product rule.

It states (as earlier) that the probability that both X and Y are true is equal to the probability that X is true given that Y is true times the probability that Y is true (independent of X).

Note all the probabilities are conditional on proposition(s) or conditioning(s) I, which denotes the relevant background information on hand.

It is important to understand that there is no such thing as an absolute probability (one without prior information).

Baye's Theorem and Marginalization

As before, we can use the sum and product rules to derive other results.

First, starting with the product rule we have

$$prob(X \cap Y|I) = prob(X|Y \cap I) \times prob(Y|I)$$

We can rewrite this equation with X and Y interchanged

$$prob(Y \cap X|I) = prob(Y|X \cap I) \times prob(X|I)$$

Since the probability that both *X* and *Y* are true must be logically the same as the probability that both *Y* and *X* are true we must also have

$$prob(Y \cap X|I) = prob(X \cap Y|I)$$

or

$$prob(X|Y \cap I) \times prob(Y|I) = prob(Y|X \cap I) \times prob(X|I)$$

or

$$prob(X|Y \cap I) = \frac{prob(Y|X \cap I) \times prob(X|I)}{prob(Y|I)}$$

which is Bayes theorem (as derived earlier).

Most standard treatments of probability do not attach much importance to Bayes' rule.

This rule, which relates prob($A|B\cap C$) to prob($B|A\cap C$), allows us to turn things around with respect to the conditioning symbol, which leads to a reorientation of our thinking about probability.

The fundamental importance of this property to data analysis becomes apparent if we replace *A* and *B* by **hypothesis** and **data**:

$$prob(A|B \cap C) \propto prob(B|A \cap C) \times prob(A|C)$$

$$prob(hypothesis|data \cap C) \propto prob(data|hypothesis \cap C) \times prob(hypothesis|C)$$

Note that the equality has been replaced with a proportionality because the term prob(data|*I*) = evidence has been omitted.

The proportionality constant can be found from the normalization requirement that the sum of the probabilities for something happening must equal 1.

The power of Bayes' theorem lies in the fact that it relates the quantity of interest, the probability that the hypothesis is true given the data, to the term that we have a better chance of being able to assign, the probability that we would have obtained the measured data if the hypothesis was true

The various terms in Bayes' theorem have formal names.

> **prob(hypothesis|C) = prior probability** represents our state of knowledge(or ignorance) about the truth of the hypothesis before we have analyzed the current data.
>
> This is modified by the experimental measurements through the term
>
> **prob(data|hypothesis ∩ C) = likelihood function**.
>
> This product gives **prob(hypothesis|data ∩ C) = posterior probability** representing our state of knowledge about the truth of the hypothesis in the light of the data(after measurements).

In some sense, Bayes' theorem encapsulates the process of learning, as we shall see later.

The book extends this discussion to the concept of **Marginalization**.

Probability Interpretation

In the standard way of thinking about probability in relation to experiments, measured results are related to probabilities using the concept of a **limit frequency**.

The limit frequency is linked to probability by this definition:

> If C can lead to either A or $\sim A$ and if in n repetitions, A occurs m times, then
>
> $$P(A|C) = \lim_{n\to\infty} \frac{m}{n}$$

We must now connect the mathematical formalism with this limit frequency concept so that we can use the formalism to make predictions for experiments in real physical systems.

This approach depends on whether we can prove that the limit makes sense for real physical systems.

Let us see how we can understand the real meaning of the above interpretation of probability and thus learn how to use it in quantum mechanics, where probability will be the dominant property.

Suppose that we have an experimental measurement, M, that can yield either A or $\sim A$ as results, with a probability for result A given by

$$P(A|M) = p$$

In general, we let any sequence of n independent measurements be labeled as event M^n and we define n_A as the number of times A occurs, where $0 \le n_A \le n$.

Now imagine we carry out a sequence of n independent measurements and we find that A occurs r times.

The probability for a sequence of results that includes result A occurring r times and $\sim A$ occurring $(n-r)$ times (independent of their order in the sequence) is given by

$$p^r q^{n-r}$$

where

$$q = P(\sim A|M) = 1 - P(A|M) = 1 - p$$

The different sequence orderings are mutually exclusive events and thus we have

$$P(n_A = r|M^n) = \sum_{all\ possible\ orderings} p^r q^{n-r}$$

The sum

$$\sum_{all\ possible\ orderings}$$

just counts the number of ways to distribute r occurrences of A and $(n\text{-}r)$ occurrences of $\sim A$, where all the terms contain the common factor $p^r q^{n-r}$.

This result is given by the Binomial probability distribution as

$$\frac{n!}{r!(n-r)!}$$

so that

$$P(n_A = r|M^n) = \frac{n!}{r!(n-r)!} p^r q^{n-r}$$

Now to get to the heart of the problem.

The frequency of A in M^n is given by

$$f_n = \frac{n_A}{n}$$

This is **not necessarily** = p in any set of measurements.

What is the relationship between them?

Consider the following:

$$\begin{aligned}\langle n_A \rangle &= \text{ average or expectation value} \\ &= \text{ sum over [possible values times probability of that value} \\ &= \sum_{r=0}^{n} r P(n_A = r|M^n) = \sum_{r=0}^{n} r \frac{n!}{r!(n-r)!} p^r q^{n-r}\end{aligned}$$

We now use a clever mathematical trick to evaluate this sum.

For the moment consider p and q to be two arbitrary independent variables.

At the end of the calculation we will let $q = 1 - p$ as is appropriate for a real physical system.

From the Binomial expansion formula, we have, in general,

$$\sum_{r=0}^{n} \frac{n!}{r!(n-r)!} p^r q^{n-r} = (p+q)^n$$

Algebra then gives

$$p\frac{\partial}{\partial p}\sum_{r=0}^{n}\frac{n!}{r!(n-r)!}p^r q^{n-r} = p\frac{\partial}{\partial p}(p+q)^n$$

$$\sum_{r=0}^{n} r\frac{n!}{r!(n-r)!}p^r q^{n-r} = np(p+q)^{n-1}$$

$$\sum_{r=0}^{n} rP(n_A = r|M^n) = np(p+q)^{n-1}$$

$$\langle n_A \rangle = np(p+q)^{n-1}$$

In a real physical system, we must have $p + q = 1$, so that we end up with the result

$$\langle n_A \rangle) = np$$

and

$$\langle f_n \rangle = \frac{\langle n_A \rangle}{n} = p$$

This says that p = the average frequency.

This does not say, however, that f_n is close to p.

More general derivation in the book.

It is important to remember:

From probability theory one derives only statements of probability, not of necessity.

First hints of "subversive" or "Bayesian" thinking.....

How do we reason in situations where it is not possible to argue with certainty?

In other words, is there a way to use the techniques of deductive logic to study the inference problem arising when using inductive logic?

No matter what scientists say, this is what they are actually doing most of the time.

The answer to this last question resides in the Bayes' rule.

To Bayes(along with Bernoulli and Laplace), a probability represented a "degree- of-belief" or "plausibility", that is, how much one thinks that something is true, **based on the evidence on hand**.

The developers of standard probability theory(Fisher, Neyman and Pearson) thought this seemed too vague and subjective a set of ideas to be the basis of a "rigorous" mathematical theory.

Therefore, they defined probability as the long-run relative frequency with which an event occurred, given infinitely many repeated experimental trials.

Since such probabilities can be measured, probability was then thought to be an objective tool for dealing with random phenomena.

Example:

Consider an urn that contains 5 red balls and 7 green balls.

If a ball is selected at random, then we would all agree that the probability of picking a red ball would be 5/12 and of picking a green ball would be 7/12.

If the ball is not returned to the urn, then it seems reasonable that the probability of picking a red or green ball must depend on the outcome of the first pick (because there will be one less red or green ball in the urn).

Now suppose that we are not told the outcome of the first pick, but are given the result of the second pick.

Does the probability of the first pick being red or green change with the knowledge of the second pick?

Initially, many observers would probably say **no**, that is, at the time of the first draw, there were still 5 red balls and 7 green balls in the urn, so the probabilities for picking red and green should still be 5/12 and 7/12 independent of the outcome of the second pick.

The error in this argument becomes clear if we consider the extreme example of an urn containing only 1 red and 1 green ball.

Although, the second pick cannot affect the first pick in a physical sense, a knowledge of the second result does influence what we can infer about the outcome of the first pick, that is, if the second ball was green, then the first ball must have been red, and vice versa.

We can calculate the result as shown below:

$$Y = \text{pick is GREEN } (2^{nd} \text{ pick})$$

$$X = \text{pick is RED } (1^{st} \text{ pick})$$

$$I = \text{initial number of RED/GREEN balls} = (n, m)$$

Non-Bayesian says:

$$prob(X|\{5,7\}) = \frac{5}{12} = 0.417$$

A Bayesian would say:

$$prob(X|Y \cap I) = \frac{prob(Y|X \cap I) \times prob(X|I)}{prob(Y|I)}$$

$$prob(X|Y \cap \{n,m\}) = \frac{prob(Y|X \cap \{n,m\}) \times \frac{n}{n+m}}{\frac{n}{n+m}\frac{m}{n+m-1} + \frac{m}{n+m}\frac{m-1}{n+m-1}}$$
$$= \frac{\frac{m}{n+m-1} \times n}{\frac{nm}{n+m-1} + \frac{m(m-1)}{n+m-1}} = \frac{n}{n+m-1}$$

$$n = m = 1 \Rightarrow prob(X|Y \cap \{1,1\}) = \frac{1}{1+1-1} = 1$$

$$n = 5, m = 7 \Rightarrow prob(X|Y \cap \{5,7\}) = \frac{5}{5+7-1} = \frac{5}{11} = 0.456$$

Clearly, the Bayesian and Non-Bayesian disagree.

However, the non-Bayesian is just **assuming** that the calculated result 0.417 is correct, whereas, the Bayesian is using the rules of probability (Bayes' Rule) to infer the result 0.456 **correctly**.

The concerns about the subjectivity of the Bayesian view of probability are understandable.

I think that the presumed shortcomings of the Bayesian approach merely reflect a confusion between subjectivity and the difficult technical question of how probabilities(especially prior probabilities) should be assigned.

The popular argument is that if a probability represents a degree-of-belief, then it must be subjective, because my belief could be different from yours.

The Bayesian view is that a probability does indeed represent how much we believe that something is true, but that this belief should be based on all the relevant information available(all prior probabilities).

While this makes the assignment of probabilities an open-ended question, because the information available to me may not be the same as that available to you, it is not the same as subjectivity.

It simply means that probabilities are always conditional, and this conditioning must be stated explicitly.

Objectivity demands only that two people having the same information should assign the same probability.

Other examples and how Bayes rule works in a learning situation are in the book.

Also in the book, I discuss a most difficult question.

How do we assign probabilities based on prior information?

Finally, in the book, I discuss Testable Information: **The Principle of Maximum Entropy**

Finally, just some general thoughts……..

In discussions of Bayesian methods, opponents often use the words subjective probabilities to say that the methods are not as valid as normal objective probability theory.

These opponents are misguided.

The main point of concern centers around the choice of the prior probability density functions (pdfs), that is, what should we do if it is not known?

This is actually a very strange question.

It is usually posed this way by opponents of the Bayesian methods in an attempt to prove its subjective nature.

No probability, whether prior, likelihood or whatever, is ever known.

It is simply an assignment which reflects the relevant information that is available.

Thus, prob($x|I_1$) ≠ prob($x|I_2$), in general, where the conditioning statements I_1 and I_2 are different.

Nevertheless, objectivity can, and must, be introduced by demanding the two people with the same information I should assign the same pdf.

I think that this consistency requirement is the most important idea of all.

While we may yet be far from knowing how to convert every piece of vague information into a concrete probability assignment, we can deal with a wide variety of problems with these ideas.

The important point is that nowhere in our discussion have we explicitly differentiated between a prior and a likelihood.

We have only considered how to assign prob($X|I$) for different types of I.

If X pertains to data, then we call prob($X|I$) a likelihood.

If neither X nor I refers to (new) measurements, then we may say it is a prior.

The distinction between the two cases is one of nomenclature and not of objectivity or subjectivity.

If it appears otherwise, then this is because we are usually prepared to state conditioning assumptions for the likelihood function but shy away from doing likewise for the prior pdf.

The use of Bayesian methods in quantum mechanics presents a very different view of quantum probability than normally appears in quantum theory textbooks.

It is becoming increasingly important in discussions of measurement.

Chapter 4

The Formulation of Quantum Mechanics

Introduction

A **physical theory** is

1. a set of basic physical concepts
2. a mathematical formalism
3. a set of connections between the physical concepts and the mathematical objects that represent them

The **process** of doing theoretical physics involves:

1. constructing a mathematical formalism that allows us to express a physical problem or experiment in terms of mathematical objects
2. solving the mathematical system representing the physical problem – this a pure mathematics at this point
3. translating the mathematical solution to the physical world using the set of connections

A **description** of a physical system will involve three ingredients:

1. variables or measurable quantities that characterize the system
2. states that describe values of the variables as a function of time
3. equations of motion that determine how the states variables change in time

In classical mechanics, the position of a particle, which is a physical concept, is connected to a set of real numbers, which is a mathematical object.

This does not, in general, present us with any conceptual difficulties because we are familiar with both positions and numbers from our everyday experiences.

In quantum mechanics, however, the mathematical formalism that we will be discussing, is very abstract.

In addition, we lack any intuition based on experience concerning quantum phenomena.

The everyday world that we live in does not appear to consist of Hermitian operators and infinite-dimensional vectors.

Throughout the last century, many experiments have shown that the various dynamical variables, which seem to have a continuous spectrum in classical mechanics, generally have completely discrete or quantized spectra or, at least, a spectrum that is both discrete and continuous.

In quantum mechanics, this property will lead to most of the so-called non-classical results and force us to work in a world of probabilities.

We will take an approach to the theoretical formalism of quantum mechanics presented in terms of postulates.

Two Postulates:
Presentation, Rationalization and Meaning

Postulate 1

> *For each dynamical variable or observable, which is a physical concept, there corresponds a Hermitian, linear operator, which is a mathematical object.*
>
> *The possible values of any measurement of the observable are restricted to the eigenvalues of the corresponding operator.*

It is the nature of a postulate that we can only make ourselves feel good about it or, in other words, make ourselves feel that it makes sense in some way.

A priori, we cannot justify it in any way and we certainly cannot prove it is true or we would have assumed a more fundamental postulate(the one we used to prove the truth).

A posteriori, we justify postulates by their effect on our predictions about experimental results.

My philosophy of theoretical physics is based on this statement

> *If the predictions agree with experiment, on certain quantum systems, then the postulates are valid for that class of systems.*

So what can we say about the 1st postulate?

We already know that linear operators possess both discrete and continuous spectra.

In addition, we have a vast storehouse of mathematical knowledge available about linear operators.

So making this choice certainly is a sensible(and the simplest) place to start.

Note that the postulate does not give us any rules for assigning operators to observables.

Singling out the eigenvalues is also sensible since they represent a special connection to the properties of the associated linear operator.

Choosing Hermitian operators also makes sense since this guarantees that we have only real eigenvalues representing measured quantities.

Now we need to deal with states.

The mathematical object that we connect to a physical state must allow us to calculate the probability distributions for all observables.

Before stating the 2nd postulate, we need to present some new mathematical objects and ideas.

$$Tr\hat{W} = 1 = \sum_k W_{kk} = \sum_k \langle \phi_k | \hat{W} | \phi_k \rangle$$

where W_{kk} is the diagonal matrix element(in the basis) of the **density operator** $\hat{W}$.

Some Properties of the Trace:

$$Tr(AB) = Tr(BA)$$

$$Tr(cB) = cTr(B)$$

$$Tr(c(A+B)) = Tr(cA) + Tr(cB) = cTr(A) + cTr(B)$$

If we denote the eigenvalues of $\hat{W}$ by w_k and the corresponding eigenvectors by $|w_k\rangle$ so that

$$\hat{W} |w_k\rangle = w_k |w_k\rangle$$

then, since $\hat{W}$ has a pure point spectrum, we can write $\hat{W}$ in terms of its eigenvalues and eigenvectors (spectral representation) as

$$\hat{W} = \sum_k w_k \left| w_k \right\rangle \left\langle w_k \right|$$

Since $\hat{W}$ is self-adjoint(Hermitian), its eigenvectors form an orthonormal basis where

$$\left\langle w_k \middle| w_j \right\rangle = \delta_{kj}$$

We now derive some properties of this density operator mathematical object.

The spectrum of $\hat{W}$ is the set of numbers $\{w_k\}$.

We then have

$$Tr\hat{W} = 1 = \sum_j \left\langle w_j \middle| \hat{W} \middle| w_j \right\rangle$$

$$= \sum_j \left\langle w_j \middle| w_j \middle| w_j \right\rangle = \sum_j w_j \left\langle w_j \middle| w_j \right\rangle = \sum_j w_j$$

So,

$$\sum_j w_j = 1$$

$\hat{W}$ is a bounded, Hermitian(self-adjoint) operator.

Since $\hat{W}$ is self-adjoint, this means that $\hat{W} = \hat{W}^\dagger$, which implies that the eigenvalues are real, i.e., $w_k = w_k^*$.

Using the fact that $\hat{W}$ is defined to be a positive operator, we then have

$$\left\langle a \middle| \hat{W} \middle| a \right\rangle = \left\langle a \middle| \left(\sum_k w_k \left| w_k \right\rangle \left\langle w_k \right| \right) \middle| a \right\rangle = \sum_k w_k \left\langle a \middle| w_k \right\rangle \left\langle w_k \middle| a \right\rangle$$

$$= \sum_k w_k \left| \left\langle a \middle| w_k \right\rangle \right|^2 \geq 0$$

for any vector $|a\rangle$.

This can only be true, in general, if

$$w_k \geq 0 \text{ for all } k$$

The results $w_k \geq 0$ for all k and $\sum_j w_j = 1$ $\longrightarrow$ $0 \leq w_k \leq 1$

Finally, for any other operator $\hat{B}$

$$\hat{W}\hat{B} = \sum_k w_k |w_k\rangle \langle w_k| \hat{B}$$

$$Tr(\hat{W}\hat{B}) = \sum_j \langle w_j| \left(\sum_k w_k |w_k\rangle \langle w_k| \hat{B} \right) |w_j\rangle$$

$$= \sum_k w_k \sum_j \langle w_j | w_k\rangle \langle w_k| \hat{B} |w_j\rangle = \sum_k w_k \sum_j \delta_{jk} \langle w_k| \hat{B} |w_j\rangle$$

$$= \sum_k w_k \langle w_k| \hat{B} |w_k\rangle$$

which is a weighted sum (where the eigenvalues of the density operator are the weight factors) of the diagonal matrix elements of the operator $\hat{B}$ in a basis consisting of eigenvectors of $\hat{W}$.

Before we can show how this connects to probabilities, we must discuss the concept of a state and its representation as a mathematical object.

In classical mechanics, a state refers to the set of coordinates and momenta of the particles in the system.

In quantum mechanics, we say that a state describes the values of a set of measurable quantities.

The state description must be probabilistic so that we can only know the probabilities that these values might occur.

In essence, we assume (this is the standard approach to quantum mechanics) that the quantum state description refers to an ensemble of similarly prepared systems(so that we can make probability statements) and NOT to any individual system in the ensemble. We note that there is some dispute about this point.

We are, thus, identifying the state directly with a set of probability distributions.

Our earlier discussion of the experimental limit frequency and its connection to probability indicated that if we know the average (or mean) value or the expectation value, which has the symbol $\langle ... \rangle$, for all of the observables in our physical system, we have then specified the state of the system as exactly as is possible.

Some properties that we require of expectation values in order that they make physical sense as average values are:

1. if $\hat{B}$ is self-adjoint, then $\langle \hat{B} \rangle$ is real
2. if $\hat{B}$ is positive, then $\langle \hat{B} \rangle$ is nonnegative
3. if $c =$ complex number, then $\langle c\hat{B} \rangle = c\langle \hat{B} \rangle$
4. $\langle \hat{A} + \hat{B} \rangle = \langle \hat{A} \rangle + \langle \hat{B} \rangle$
5. $\langle \hat{I} \rangle = 1$

Postulate 2

(a) *A density operator exists for every real physical system.*

This rather innocuous looking postulate is at the heart of quantum theory.

Its full meaning will only become clear as we learn its connection to the probability interpretation of quantum mechanics.

(b) *The expectation value of an operator* $\hat{B}$ *is given by*

$$\langle \hat{B} \rangle = Tr(\hat{W}\hat{B})$$

If we assume that every bounded, Hermitian operator represents some measurable quantity for a physical system, then each state is represented by a unique density operator $\hat{W}$.

Let us choose a simple example of a density operator to get some handle on what this postulate is saying.

In particular, let us choose as our density operator $\hat{W}$ the projection operator onto a 1-dimensional space spanned by the vector $|\alpha\rangle$, namelv

$$\hat{W} = |\alpha\rangle\langle\alpha|$$

Since $\hat{W}^2 = \hat{W}$ (= an idempotent operator), it has eigenvalues $w_k = 0,1$ only.

The eigenvector corresponding to eigenvalue 1 is $|\alpha\rangle$.

We also have

$$\sum_k w_k = 1 = Tr(\hat{W})$$

$$\langle a|\hat{W}|a\rangle = |\langle a|\alpha\rangle|^2 \geq 0$$

so that all required properties for being a density operator are satisfied.

Since,

$$0 \le w_k \le 1 \text{ and } \sum_k w_k = 1$$

we can think of

$$\hat{W} = \sum_k w_k \left|w_k\right\rangle \left\langle w_k\right|$$

as representing a **statistical mixture** of the states represented by the vectors $|w_k\rangle$ where the probability is w_k that we have the state $|w_k\rangle$ present.

In this simple case, we then have

$$\langle \hat{B} \rangle = Tr(\hat{W}\hat{B}) = \langle \alpha | \hat{B} | \alpha \rangle$$

which is the expectation value of $\hat{B}$ in the state $|\alpha\rangle$.

Proof: Let $|w_1\rangle = |\beta\rangle$, $|w_2\rangle = |\alpha\rangle$, $\langle \alpha | \beta \rangle = 0$. Then

$$\begin{aligned} Tr(\hat{W}\hat{B}) &= \sum_{m=1}^{2} \langle w_m | \hat{W}\hat{B} | w_m \rangle = \sum_{m=1}^{2} \langle w_m | \alpha \rangle \langle \alpha | \hat{B} | w_m \rangle \\ &= \sum_{m=1}^{2} \langle w_m | w_2 \rangle \langle w_2 | \hat{B} | w_m \rangle = \sum_{m=1}^{2} \delta_{m2} \langle w_2 | \hat{B} | w_m \rangle \\ &= \langle w_2 | \hat{B} | w_2 \rangle = \langle \alpha | \hat{B} | \alpha \rangle \end{aligned}$$

Since the important quantities for connection to experiment will be expectation values, we see that the state represented by $\hat{W}$, in this case, is **equally well** represented by the state vector $|\alpha\rangle$; the density operator and the state vector are **equivalent** ways of representing a physical system **in this simple case**.

In general, when

$$\hat{W} = \sum_k w_k \left|w_k\right\rangle \left\langle w_k\right|$$

we saw earlier that

$$Tr(\hat{W}\hat{B}) = \sum_k w_k \left\langle w_k\right| \hat{B} \left|w_k\right\rangle$$

Postulate #2 then says that the average or expectation value of $\hat{W}$, $\langle \hat{B} \rangle$, equals the weighted sum of the expectation value in each of the eigenstates $|w_k\rangle$ of $\hat{W}$ with a weighting factor equal to the corresponding eigenvalue w_k.

Thus, if we know the numbers $\langle w_k| \hat{B} |w_k\rangle$ we can find the expectation value of the corresponding operator, which is the maximum amount of information we can know.

States and Probabilities

States

Among the set of all possible states there exists a special group **called pure states**.

The density operator for a pure state is given by

$$\hat{W} = \left|\psi\right\rangle \left\langle\psi\right|$$

where the vector $|\psi\rangle$ is of unit norm and is called the state vector.

We discussed this density operator above and found that the expectation value for an observable represented by $\hat{Q}$ now in this pure state $|\psi\rangle$, is given by

$$\langle\hat{Q}\rangle = Tr(\hat{W}\hat{Q}) = Tr(\left|\psi\right\rangle \left\langle\psi\right| \hat{Q}) = \left\langle\psi\right| \hat{Q} \left|\psi\right\rangle$$

Another property of a pure state goes as follows: we have, in the general case,

$$0 \leq w_k \leq 1 \text{ and } \sum_k w_k = 1$$

which implies that

$$w_k^2 \leq w_k$$

$$Tr(\hat{W}^2) = \sum_k w_k^2 \leq \sum_k w_k = 1$$

$$Tr(\hat{W}^2) \leq 1$$

For a pure state, however, where $w_k{}^2 = w_k$ because $\hat{W}$ is an idempotent projection operator, we then have

$$Tr(\hat{W}^2) = 1$$

The most important way of distinguishing whether a state is pure or not follows from this property of density operators:

> *The density operator for a pure state cannot be written as a linear combination of the density operators of other states, but the density operator for a nonpure state can always be so written.*

We will assume that both **pure** and **nonpure**(other texts use the name **mixed state** or **mixture**) states are fundamental and we will see physical systems later on that are described by each type.

Probabilities

Postulate #2 says that the average value of an observable represented by $\hat{Q}$ in the state corresponding to density operator $\hat{W}$ is given by

$$\langle \hat{Q} \rangle = Tr(\hat{W}\hat{Q})$$

Consider a function F($\hat{Q}$) of the operator $\hat{Q}$.

Let

$$h(q)\,dq = \text{probability that the measured value of the observable represented by } \hat{Q} \text{ lies between } q \text{ and } q + dq$$

Thus,

$$h(q) = \text{probability density function}$$

Now, the general definition of an average value says that

$$\langle F(\hat{Q}) \rangle = \int_{-\infty}^{\infty} F(q')h(q')\,dq'$$

Postulate #2, however, says

$$\langle F(\hat{Q}) \rangle = Tr(\hat{W}F(\hat{Q}))$$

Taken together, these two relations will allow us to extract the probability density function $h(q)$.

Case 1 : Discrete Spectrum

Let $\hat{Q}$ be a self-adjoint operator representing the observable Q.

Assume that it has a pure discrete spectrum with eigenvectors $|q_n\rangle$ and corresponding eigenvalues q_n such that

$$\hat{Q}\,|q_n\rangle = q_n\,|q_n\rangle$$

We can write its spectral representation (express the operator in terms of its eigenvectors and eigenvalues) as

$$\hat{Q} = \sum_n q_n\,|q_n\rangle\,\langle q_n|$$

Now consider the function

$$F(Q) = \theta(q - Q)$$

where

$$\theta(q - Q) = \begin{cases} 1 & q > Q \\ 0 & q < Q \end{cases}$$

The expectation value of this function is then

$$\langle F(\hat{Q})\rangle = \langle \theta(q - \hat{Q})\rangle = \int_{-\infty}^{\infty} F(q')h(q')\,dq' = \int_{-\infty}^{\infty} \theta(q - q')h(q')\,dq'$$

$$= \int_{-\infty}^{q} h(q')\,dq' = Prob(Q < q|\hat{W})$$

$$= \text{probability that } Q < q \text{ given } \hat{W}$$

However, we can also write

$$\langle\theta(q-\hat{Q})\rangle = Tr(\hat{W}\theta(q-\hat{Q})) = Tr\left[\hat{W}\sum_n \theta(q-q_n)\,|q_n\rangle\langle q_n|\right]$$

$$= \sum_m \langle q_m|\left(\hat{W}\sum_n \theta(q-q_n)\,|q_n\rangle\langle q_n|\right)|q_m\rangle$$

$$= \sum_m\sum_n \langle q_m|\,\hat{W}\,|q_n\rangle\,\theta(q-q_n)\,\langle q_n\,|\,q_m\rangle$$

$$= \sum_m\sum_n \langle q_m|\,\hat{W}\,|q_n\rangle\,\theta(q-q_n)\delta_{nm} = \sum_n \langle q_n|\,\hat{W}\,|q_n\rangle\,\theta(q-q_n)$$

$$= Prob(Q<q|\hat{W})$$

Now we have

$$h(q) = \frac{\partial}{\partial q}\int_{-\infty}^{q} h(q')\,dq' = \frac{\partial}{\partial q}Prob(Q<q|\hat{W})$$

$$= \frac{\partial}{\partial q}\sum_n \langle q_n|\,\hat{W}\,|q_n\rangle\,\theta(q-q_n) = \sum_n \langle q_n|\,\hat{W}\,|q_n\rangle\,\delta(q-q_n)$$

This implies that the probability density $h(q) = 0$ unless $q =$ an eigenvalue.

In other words, the probability = 0 that a dynamical variable or observable will take on a value other than an eigenvalue of the corresponding operator.

This makes it clear the postulate #2 is consistent with the last part of postulate #1, where we assumed this to be the case.

Now

$Prob(Q = q|\hat{W})$ = probability that the observable represented by $\hat{Q}$ will have the discrete value q in the ensemble characterized by $\hat{W}$

We calculate this as follows:

$$Prob(Q = q|\hat{W}) = \lim_{\epsilon\to 0}(Prob(Q < q+\epsilon|\hat{W}) - Prob(Q < q-\epsilon|\hat{W}))$$

$$= \lim_{\epsilon\to 0}\left(\sum_n \langle q_n|\hat{W}|q_n\rangle\,\theta(q+\epsilon-q_n) - \sum_n \langle q_n|\hat{W}|q_n\rangle\,\theta(q-\epsilon-q_n)\right)$$

$$= \sum_n \langle q_n|\hat{W}|q_n\rangle \lim_{\epsilon\to 0}(\theta(q+\epsilon+q_n) - \theta(q+\epsilon-q_n))$$

$$= \sum_n \langle q_n|\hat{W}|q_n\rangle \lim_{\epsilon\to 0}(\delta(q-q_n)2\epsilon) = \sum_n \langle q_n|\hat{W}|q_n\rangle\,\delta_{q,q_n}$$

If we now define the projection operator

$$\hat{P}(q) = \sum_n |q_n\rangle\langle q_n|\,\delta_{q,q_n}$$

which projects onto the subspace spanned by all the eigenvectors (degenerate) with eigenvalue $q = q_n$, we then have the general result that

$$Tr(\hat{W}\hat{P}(q)) = \sum_m \langle q_m|\hat{W}\hat{P}(q)|q_m\rangle$$

$$= \sum_m \langle q_m|\hat{W}\left(\sum_n |q_n\rangle\langle q_n|\,\delta_{q,q_n}\right)|q_m\rangle$$

$$= \sum_{m,n} \langle q_m|\hat{W}|q_n\rangle\,\delta_{q,q_n}\,\langle q_n|q_m\rangle$$

$$= \sum_{m,n} \langle q_m|\hat{W}|q_n\rangle\,\delta_{q,q_n}\delta_{nm}$$

$$= \sum_n \langle q_n|\hat{W}|q_n\rangle\,\delta_{q,q_n} = Prob(Q = q|\hat{W})$$

If we have a pure state, then

$$\hat{W} = |\psi\rangle\langle\psi|$$

and if q_n is a non-degenerate eigenvalue,

$$\begin{aligned} Prob(Q = q_n|\hat{W}) &= Tr(\hat{W}\hat{P}(q_n)) = Tr((|\psi\rangle\langle\psi|)\hat{P}(q_n)) \\ &= \sum_m \langle q_m|(|\psi\rangle\langle\psi|)\hat{P}(q_n)|q_m\rangle \\ &= \sum_m \langle q_m|(|\psi\rangle\langle\psi|)(|q_n\rangle\langle q_n|)|q_m\rangle \\ &= \sum_m \langle q_m|\psi\rangle\langle\psi|q_n\rangle\langle q_n|q_m\rangle \\ &= \sum_m \langle q_m|\psi\rangle\langle\psi|q_n\rangle\delta_{nm} \\ &= \langle q_n|\psi\rangle\langle\psi|q_n\rangle = |\langle q_n|\psi\rangle|^2 \end{aligned}$$

We note the following alternate and useful form of this result.

We have

$$\begin{aligned} Prob(Q = q_n|\hat{W}) &= |\langle q_n|\psi\rangle|^2 = \langle q_n|\psi\rangle\langle\psi|q_n\rangle \\ &= \langle q_n|\hat{W}|q_n\rangle = \sum_m \langle q_m|\hat{P}_{q_n}\hat{W}|q_m\rangle \\ &= Tr(\hat{P}_{q_n}\hat{W}) \end{aligned}$$

Thus, for a physical system that is characterized by a density operator $\hat{W} = |\psi\rangle\langle\psi|$ or represented by a pure state characterized by a state vector $|\psi\rangle$, the probability of measuring a particular eigenvalue q_n of an observable Q represented by the operator $\hat{Q}$ is given by

$$|\langle q_n|\psi\rangle|^2$$

To see that this all makes sense relative to the standard definition of an average value we consider the quantity

$$\langle\alpha|\hat{Q}|\alpha\rangle = \text{ average value in the state } |\alpha\rangle$$

Now using the eigenvector/eigenvalue relation

$$\hat{Q}|q_k\rangle = q_k|q_k\rangle$$

which says that we can write

$$\hat{Q} = \sum_k q_k |q_k\rangle\langle q_k|$$

we get

$$\langle\alpha|\hat{Q}|\alpha\rangle = \sum q_k \langle\alpha|q_k\rangle\langle q_k|\alpha\rangle = \sum q_k |\langle q_k|\alpha\rangle|^2$$

which says that the average value in the state $|\alpha\rangle$ is equal to the sum over all possible values (the eigenvalues) of the value(eigenvalue) times the probability of finding measuring that eigenvalue when in the state $|\alpha\rangle$, which is clearly given by

$$|\langle q_k|\alpha\rangle|^2$$

This corresponds to the standard definition of the average value.

Thus, the connection between the mathematical object $|\langle q_k | \alpha\rangle|^2$ and the probability of measuring the eigenvalue q_n when in the state $|\alpha\rangle$ is now clear and derivable from the earlier postulates.

We do not need to assume this as another postulate as is done in many texts.

In the book I show that the only state for which the observable takes on the value q_0 (non-degenerate) with probability = 1 is the pure state represented by the density operator $\hat{W} = |q_0\rangle\langle q_0|$.

This state, whether described by the density operator $\hat{W}$ or the state vector $|q_0\rangle$ is called an eigenstate of $\hat{Q}$.

The observable has a definite value (probability = 1) in an eigenstate and ONLY in an eigenstate.

Case 2: Continuous Spectrum

The full-blown mathematics for the continuous spectrum is done in Chapter 2 of the book.

Here we will just assume that variables are continuous and sums become integrals.

Now we let $\hat{Q}$ be a self-adjoint operator with a purely continuous spectrum, so that we can write

$$\hat{Q} = \int q' \, |q'\rangle \langle q'| \, dq' \text{ where } \hat{Q} \, |q'\rangle = q' \, |q'\rangle$$

The eigenvectors are normalized using the Dirac delta-function, as we showed in Chapter 2 of book

$$\langle q' \, | \, q'' \rangle = \delta(q' - q'')$$

and we let $h(q)\,dq =$ probability that the measured value of the observable represented by $\hat{Q}$ lies between q and $q + dq$

Then, as in the discrete case, we have

$$\langle \theta(q - \hat{Q}) \rangle = \int_{-\infty}^{q} h(q')\,dq' = Prob(Q < q|\hat{W})$$

which gives the probability that Q < q.

We also have

$$\langle \theta(q-\hat{Q}) \rangle = Tr(\hat{W}\theta(q-\hat{Q}))$$

and using the standard expansion(see book Chapter 2) of an operator

$$\theta(q-\hat{Q}) = \int_{-\infty}^{\infty} \theta(q-q')\,|q'\rangle\,\langle q'|\,dq'$$

we get

$$\begin{aligned}
\theta(q-\hat{Q}) &= Tr\left(\hat{W}\int_{-\infty}^{\infty}\theta(q-q')\,|q'\rangle\,\langle q'|\,dq'\right) \\
&= \int_{-\infty}^{\infty} dq''\,\langle q''|\,\hat{W}\left(\int_{-\infty}^{\infty}\theta(q-q')\,|q'\rangle\,\langle q'||q''\rangle\,dq'\right) \\
&= \int_{-\infty}^{\infty} dq''\,\langle q''|\,\hat{W}\int_{-\infty}^{\infty}\theta(q-q')\,|q'\rangle\,\langle q'\,|q''\rangle\,dq' \\
&= \int_{-\infty}^{\infty} dq''\,\langle q''|\,\hat{W}\int_{-\infty}^{\infty}\theta(q-q')\,|q'\rangle\,\delta(q'-q'')\,dq' \\
&= \int_{-\infty}^{\infty}\theta(q-q')\,\langle q'|\,\hat{W}\,|q'\rangle\,dq' \\
&= \int_{-\infty}^{q}\langle q'|\,\hat{W}\,|q'\rangle\,dq' = Prob(Q<q|\hat{W})
\end{aligned}$$

This says that the probability density for the observable represented by $\hat{Q}$ within the ensemble characterized by $\hat{W}$ is

$$h(q) = \frac{\partial}{\partial q}Prob(Q<q|\hat{W}) = \langle q|\,\hat{W}\,|q\rangle$$

in general.

For the special case of a pure state where $\hat{W} = |\psi\rangle\langle\psi|$ we have

$$h(q) = |\langle q|\psi\rangle|^2$$

Later on we shall find that this is equivalent to a probability statement for the famous Schrodinger equation wave function.

We now turn to the topic of quantum dynamics.

As we have seen :

States or density operators describe values or probabilities for measurable quantities at given times.

We want to find :

Equations of motion that determine how these values or probabilities of measurable quantities change with time.

It is clear from our discussion, that, in the theory of quantum mechanics, a state only needs to be able specify how to calculate expectations values of measurable quantities.

Therefore, we will assume that

Equations of motion, at a minimum, only need to specify how expectation values change in time.

Quantum Pictures

There are three ways that are commonly used in quantum mechanics to make expectation values depend on time:

Schrodinger Picture

1. states are represented by ket vectors that depend on time, $|\psi(t)\rangle$
2. operators representing observables or measurable quantities are independent of time, $\hat{Q}$

We then get a time-dependent expectation value of the form

$$\langle \hat{Q}(t) \rangle = \langle \psi(t) | \hat{Q} | \psi(t) \rangle$$

Heisenberg Picture

1. operators representing observables or measurable quantities are dependent of time, $\hat{Q}(t)$
2. states are represented by ket vectors that do not depend on time, $|\psi\rangle$

We then get a time-dependent expectation value of the form

$$\langle \hat{Q}(t) \rangle = \langle \psi | \hat{Q}(t) | \psi \rangle$$

Interaction Picture

It is a mixture of the Schrodinger and Heisenberg pictures that is appropriate for a very important class of problems.

We will discuss it later after presenting the Schrodinger and Heisenberg pictures.

All of these pictures must agree in the sense that they must all give the same time dependence for $\langle \hat{Q}(t) \rangle$.

There is, after all, a unique real world out there!!!

We will discuss these **pictures** in terms of state vectors first, mixing in state operator aspects as we proceed.

We have been discussing the so-called **formal structure** of quantum mechanics.

This structure is a fundamental part of the theory behind quantum mechanics, but it has very little physical content on its own.

We cannot solve any physical problem with the formalism as it stands.

We must first develop connections or correspondence rules that tell us the specific operators that actually represent particular dynamical variables.

The fundamental dynamical variables that we will be working with are **position, linear momentum, angular momentum and energy**.

All such quantities are related to space-time symmetry transformations.

As we proceed with our discussions, we will introduce any needed aspects of symmetry transformations and discuss further aspects of the subject later.

Transformations of States and Observables

The way it must be.....

Experimental evidence leads us to believe that the laws of nature are invariant under certain space-time symmetry transformations, including displacements in space and time, rotations and Lorentz boosts (relative velocity changes).

For each such symmetry, both the state vectors and the observables must have transformations, i.e.,

$$|\psi\rangle \rightarrow |\psi'\rangle \qquad , \qquad \hat{Q} \rightarrow \hat{Q}'$$

We will only use pure states for most of our discussions in this development since nothing new appears in more complex cases.

What must be preserved in these transformations?

1. If $\hat{Q}\,|q_n\rangle = q_n\,|q_n\rangle$, then $\hat{Q}'\,|q'_n\rangle = q_n\,|q'_n\rangle$

   Here we are assuming that the eigenvalues are unchanged since $\hat{Q}$ and $\hat{Q}'$ are the SAME observable represented in two frames of reference.

   Mathematically operators representing the same dynamical variable must have the same spectrum.

   For example, in both frames the position operator will have identical continuous spectra in the range $[-\infty,\infty]$ in each frame.

2. If $|\psi\rangle = \sum_n c_n\,|q_n\rangle$ where $\hat{Q}\,|q_n\rangle = q_n\,|q_n\rangle$, then $|\psi'\rangle = \sum_n c'_n\,|q'_n\rangle$ where $\hat{Q}'\,|q'_n\rangle = q_n\,|q'_n\rangle$.

   This actually follows from (1).

Now, equivalent events observed in each frame must have the same probability.

If this were not true, then some event might occur more often in one frame than it does in another frame, which makes no physical sense.

This means that probabilities are equal or that

$$|c_n|^2 = |c'_n|^2 \text{ or } |\langle q_n\,|\,\psi\rangle|^2 = |\langle q'_n\,|\,\psi'\rangle|^2$$

We now present the mathematical formalism that characterizes this type of transformation.

The book has a digression at this point to cover two new mathematical topics, antiunitaty/antilinear operators and Wigner's Theorem.

I will just use their properties in proofs, if necessary and you can refer to the book for the mathematical details.

We can show that all transformation operators of interest in quantum mechanics are linear operators.

For example, let $\hat{U}(\ell)$ describe a displacement through a distance ℓ.

We know from experiment that this can be done as a series of two displacements of size $\ell/2$ and thus, we must have $\hat{U}(\ell) = \hat{U}(\ell/2)\,\hat{U}(\ell/2)$.

Regardless of whether $\hat{U}(\ell/2)$ is linear or antilinear, $\hat{U}(\ell)$ is linear.

There is nothing special, however, about the value ℓ, i.e., we must also have $\hat{U}(\ell/2) = \hat{U}(\ell/4)\ \hat{U}(\ell/4)$ which implies that $\hat{U}(\ell/2)$ is linear and so on.

Operators in quantum mechanics are continuous, which means that they cannot change discontinuously from linear to antilinear as a function of ℓ.

This means that we need only consider continuous linear transformations in quantum mechanics.

Now, if the transformation rule for state vectors is $|q'_n\rangle = \hat{Q}\,|q_n\rangle$ where

$$\hat{Q}\,|q_n\rangle = q_n\,|q_n\rangle \text{ and } \hat{Q}'\,|q'_n\rangle = q_n\,|q'_n\rangle$$

then we must have

$$\hat{Q}'\hat{U}\,|q_n\rangle = q_n\hat{U}\,|q_n\rangle \text{ or } \hat{U}^{-1}\hat{Q}'\hat{U}\,|q_n\rangle = q_n\,|q_n\rangle$$

Therefore,

$$(\hat{Q} - \hat{U}^{-1}\hat{Q}'\hat{U})\,|q_n\rangle = (q_n - q_n)\,|q_n\rangle = 0$$

for all $|q_n\rangle$.

Since the set $\{|q_n\rangle\}$ is complete (the eigenvectors of a Hermitian operator), this result holds for any vector $|\psi\rangle$, which can be constructed from the set $\{|q_n\rangle\}$.

Therefore, we must have

$$\hat{Q} - \hat{U}^{-1}\hat{Q}'\hat{U} = 0$$

or

$$\hat{Q} \to \hat{Q}' = \hat{U}\hat{Q}'\hat{U}^{-1}$$

is the corresponding transformation rule for linear operators.

The Transformation Operator and its Generator

Let t be a continuous parameter.

We consider a family of unitary operators $\hat{U}(t)$, with the properties

$$\hat{U}(0) = \hat{I} \text{ and } \hat{U}(t_1 + t_2) = \hat{U}(t_1)\hat{U}(t_2)$$

Transformations such as displacements, rotations and Lorentz boosts clearly satisfy these properties and so it make sense to require them in general.

Now we consider infinitesimal t.

We can then write the infinitesimal version of the unitary transformation as

$$\hat{U}(t) = \hat{I} + \left.\frac{d\hat{U}}{dt}\right|_{t=0} t + O(t^2)$$

Now all unitary operators must satisfy the unitarity condition

$$\hat{U}\hat{U}^{\dagger} = \hat{I} \text{ for all } t$$

Therefore, we have to first order in t (an infinitesimal)

$$\hat{U}\hat{U}^{\dagger} = \hat{I} = \left[\hat{I} + \left.\frac{d\hat{U}(t)}{dt}\right|_{t=0} t + \ldots\right]\left[\hat{I} + \left.\frac{d\hat{U}^{\dagger}(t)}{dt}\right|_{t=0} t + \ldots\right]$$

$$= \hat{I} + \left[\frac{d\hat{U}(t)}{dt} + \frac{d\hat{U}^{\dagger}(t)}{dt}\right]\Bigg|_{t=0} t + \ldots$$

which implies the condition

$$\left[\frac{d\hat{U}(t)}{dt} + \frac{d\hat{U}^{\dagger}(t)}{dt}\right]\Bigg|_{t=0} = 0$$

If we let

$$\left.\frac{d\hat{U}(t)}{dt}\right|_{t=0} = -i\hat{H}$$

then the condition becomes

$$-i\hat{H} + (i\hat{H})^{\dagger} = 0 \text{ or } \hat{H} = \hat{H}^{\dagger}$$

which says that $\hat{H}$ is a Hermitian operator.

It is called the **generator** of the family of transformations $\hat{U}$(t) because it determines these operators uniquely.

Now consider the property $\hat{U}(t_1 + t_2) = \hat{U}(t_1)\ \hat{U}(t_2)$.

Taking the appropriate partial derivative we have

$$\frac{\partial}{\partial t_1}\hat{U}(t_1 + t_2))\Big|_{t_1=0} = \left(\frac{d}{dt_1}\hat{U}(t_1)\right)\Big|_{t_1=0}\hat{U}(t_2)$$

or

$$\frac{d}{dt}\hat{U}(t))\Big|_{t=t_2} = -i\hat{H}\hat{U}(t_2)$$

which can be written for arbitrary t as

$$i\frac{d\hat{U}(t)}{dt} = \hat{H}\hat{U}(t)$$

If $\hat{H}$ is **not explicitly dependent on time**, then this equation is satisfied by the unique solution

$$\hat{U}(t) = e^{-i\hat{H}t}$$

Thus, the generator $\hat{H}$ of the infinitesimal transformation, determines the operator $\hat{U}(t) = e^{-i\hat{H}t}$ for a finite transformation.

This is just Stone's theorem, which is discussed in the book, but now has been derived in an alternative way.

We will now approach the study of various pictures using simple methods first and then repeat the process using symmetry operations, which will clearly show the power of using the latter approach and give us a deeper understanding about what is happening.

The Schrodinger Picture

The Schrodinger picture follows directly from the previous discussion of the $\hat{U}(t)$ operator.

Suppose we have some physical system that is represented by the state vector $|\psi(0)\rangle$ at time $t = 0$ and represented by the state vector $|\psi(t)\rangle$ at time t.

We ask this question. How are these state vectors related to each other?

We make the following assumptions:

1. *Every vector* $|\psi(0)\rangle$ such that $\langle\psi(0)|\psi(0)\rangle = \| |\psi(0)\rangle \| = 1$ represents a *possible state* at time $t = 0$.

2. *Every vector* $|\psi(t)\rangle$ such that $\langle\psi(t)|\psi(t)\rangle = \| |\psi(t)\rangle \| = 1$ represents a *possible state* at time $t = 0$.

3. *Every bounded Hermitian operator* represents an *observable* or *measurable* quantity.

4. The properties of the physical system determine the state vectors to *within a phase factor* since $\| e^{i\alpha} |\psi\rangle \| = \| |\psi\rangle \|$.

5. $|\psi(t)\rangle$ is *determined* by $|\psi(0)\rangle$. Now if $|\psi(0)\rangle$ and $|\phi(0)\rangle$ represent two possible states at $t = 0$ and $|\psi(t)\rangle$ and $|\phi(t)\rangle$ represent the *corresponding* states at time t, then $|\langle\phi(0)|\psi(0)\rangle|^2$ equals the probability of finding the system in the state represented by $|\phi(0)\rangle$ given that the system is in the state $|\psi(0)\rangle$ at $t = 0$ and $|\langle\phi(t)|\psi(t)\rangle|^2$ equals the probability of finding the system in the state represented by $|\phi(t)\rangle$ given that the system is in the state $|\psi(t)\rangle$ at time t.

6. It makes physical sense to assume that these two probabilities should be the same

$$|\langle\phi(0)|\psi(0)\rangle|^2 = |\langle\phi(t)|\psi(t)\rangle|^2$$

Wigner's (see book) then says that there exists a unitary, linear operator $\hat{U}(t)$ such that

$$|\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle$$

and an expression of the form

$$|\langle\alpha|\,\hat{U}(t)\,|\beta\rangle\,|^2$$

gives the probability that the system is in state $|\alpha\rangle$ at time t given that it was in state $|\beta\rangle$ at time $t = 0$.

This clearly agrees with our earlier assumption of the existence of such an operator and strengthens our belief that item (6) above is a valid assumption.

We assume that this expression is a continuous function of t.

As we have already showed, we then have $\hat{U}(t)$ satisfying the equation

$$i\frac{d\hat{U}(t)}{dt} = \hat{H}\hat{U}(t) \qquad \text{or} \qquad \hat{U}(t) = e^{-i\hat{H}t}$$

and thus

$$|\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle = e^{-i\hat{H}t}\,|\psi(0)\rangle$$

which implies the following equation of motion for the state vector

$$i\frac{d\hat{U}(t)}{dt}\,|\psi(0)\rangle = \hat{H}\hat{U}(t)\,|\psi(0)\rangle \qquad \text{or} \qquad i\frac{d}{dt}\,|\psi(t)\rangle = \hat{H}\,|\psi(t)\rangle$$

which is the abstract form of the famous Schrodinger equation.

We will derive the standard form of this equation later.

The operator $\hat{U}(t) = e^{-i\hat{H}t}$ is called the time evolution operator for reasons that will become clear shortly.

Finally, we can write a time-dependent expectation value as

$$|\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle = e^{-i\hat{H}t}\,|\psi(0)\rangle$$

$$\langle\hat{Q}(t)\rangle = \langle\psi(t)|\,\hat{Q}\,|\psi(t)\rangle$$

This is the **Schrodinger picture** where **state vectors change with time** and **operators are constant in time**.

As we saw in the discussion above, using the Schrodinger picture depends on a full knowledge of the Hamiltonian operator .

However, in the Schrodinger picture, where we need to know $\hat{H}$ to solve the equation of motion for $|\psi(t)\rangle$, the equation of motion is such that we seem to need to know the complete solution for all time, $|\psi(t)\rangle$ to deduce $\hat{H}$.

We are trapped in a circle.

Put another way, the Schrodinger equation has no physical content unless we have an independent way to choose the Hamiltonian operator $\hat{H}$.

Before deriving a way to choose $\hat{H}$ (we will use a symmetry approach), we will look at the other pictures.

We note that the Schrodinger picture is not the same as the Schrodinger equation.

The Schrodinger equation involves a mathematical object called the wave function which is one particular representation of the state vector, namely the position representation, as we shall see later.

Thus, the Schrodinger equation is applicable only to Hamiltonians that describe operators dependent on external degrees of freedom like position and momentum.

The Schrodinger picture, on the other hand, works with both internal and external degrees of freedom and can handle a much wider class of physical systems, as we will see.

The Heisenberg Picture

As we saw earlier, we can think of the expectation value in two different ways, namely,

$$|\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle = e^{-i\hat{H}t}\,|\psi(0)\rangle$$

$$\begin{aligned}\langle\hat{Q}(t)\rangle &= \langle\psi(t)|\,\hat{Q}\,|\psi(t)\rangle = \langle\psi(0)|\,\hat{U}^{\dagger}(t)\hat{Q}\hat{U}(t)\,|\psi(0)\rangle \\ &= \langle\psi(0)|\,\hat{Q}(t)\,|\psi(0)\rangle\end{aligned}$$

where

$$\hat{Q}(t) = \hat{U}(t)^{\dagger}\hat{Q}\hat{U}(t) = \hat{U}^{\dagger}(t)\hat{Q}(0)\hat{U}(t) = e^{i\hat{H}t}\hat{Q}e^{-i\hat{H}t}$$

This implies that in the Heisenberg picture **operators change with time** and **states are constant in time**.

We will not be using the Heisenberg picture in our discussion, so more details about the Heisenberg picture can be found in the book.

Interaction Picture

We have derived the equation

time-dependent

$$i\frac{d}{dt}|\psi(t)\rangle = \hat{H}(t)|\psi(t)\rangle$$

Now let us assume that $\hat{H}$ can be split up as follows:

$$\hat{H}(t) = \hat{H}_0 + \hat{V}(t)$$

where $\hat{H}_0$ is independent of t.

This will be possible in many real physical systems.

Note that the equation

$$i\frac{d\hat{U}(t)}{dt} = \hat{H}(t)\hat{U}(t)$$

does not have the simple solution

$$\hat{U}(t) = e^{-i\hat{H}t}$$

in this case.

We now define a new state vector by the relation

$$|\psi_I(t)\rangle = e^{i\hat{H}_0 t}|\psi(t)\rangle$$

Taking derivatives we get

$$\begin{aligned} i\frac{d}{dt}|\psi_I(t)\rangle &= i\frac{d}{dt}(e^{i\hat{H}_0 t}|\psi(t)\rangle) = -e^{i\hat{H}_0 t}\hat{H}_0|\psi_I(t)\rangle + e^{i\hat{H}_0 t}i\frac{d}{dt}|\psi(t)\rangle \\ &= -e^{i\hat{H}_0 t}\hat{H}_0|\psi_I(t)\rangle + e^{i\hat{H}_0 t}\hat{H}|\psi(t)\rangle = e^{i\hat{H}_0 t}\hat{V}|\psi(t)\rangle \\ &= e^{i\hat{H}_0 t}\hat{V}e^{-i\hat{H}_0 t}e^{i\hat{H}_0 t}|\psi(t)\rangle \end{aligned}$$

or

$$i\frac{d}{dt}|\psi_I(t)\rangle = \hat{V}_I(t)\,|\psi_I(t)\rangle$$

where we have defined

$$\hat{V}_I(t) = e^{i\hat{H}_0 t}\hat{V}e^{-i\hat{H}_0 t}$$

We then have

$$\langle\hat{Q}(t)\rangle = \langle\psi(t)|\,\hat{Q}\,|\psi(t)\rangle = \langle\psi_I(t)|\,e^{i\hat{H}_0 t}\hat{Q}e^{-i\hat{H}_0 t}\,|\psi_I(t)\rangle$$
$$= \langle\psi_I(t)|\,\hat{Q}_I(t)\,|\psi_I(t)\rangle$$

This says that in the interaction picture both the state vectors and the operators are dependent on time.

Their time development, however, depends on different parts of $\hat{H}$.

The state vector time development depends on $\hat{V}_I(t)$ and the operator time development depends on $\hat{H}_0$.

It is, in some sense, intermediate between the Schrodinger and Heisenberg picture.

The question still remains, however, how do we find $\hat{H}$?

We now turn to a more general approach based on symmetries to get a handle on how to deal with this problem.

Symmetries of Space-Time

Since we are only considering non-relativistic quantum mechanics at this stage, we will restrict our attention to velocities that are small compared to the speed of light.

In this case, the set of all displacements in space-time, rotations in space and Galilean boosts(Lorentz boosts in the low velocity limit) can be represented by transformations where

$$\vec{x} \to \vec{x}' = \mathbb{R}\vec{x} + \vec{a} + \vec{v}t \quad , \quad t' = t + s$$

where $\mathbb{R}$ is a rotation of the real 3-vectors $\vec{x}$ in 3-dimensional space, $\vec{a}$ is a real 3-vector that specifies space translations, $\vec{v}$ is a real 3-vector that represents the velocity of a moving coordinate system and specifies the Galilean transformations, and the real number s specifies the time translation.

$\mathbb{R}$ can be thought of as a 3×3 matrix such that under a pure rotation

$$\vec{x}' = \mathbb{R}\vec{x} \text{ or } x'_j = \sum_i R_{ji} x_i$$

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

For example, a rotation by angle θ about the x_3(or z)-axis is given by

$$\mathbb{R}_3(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

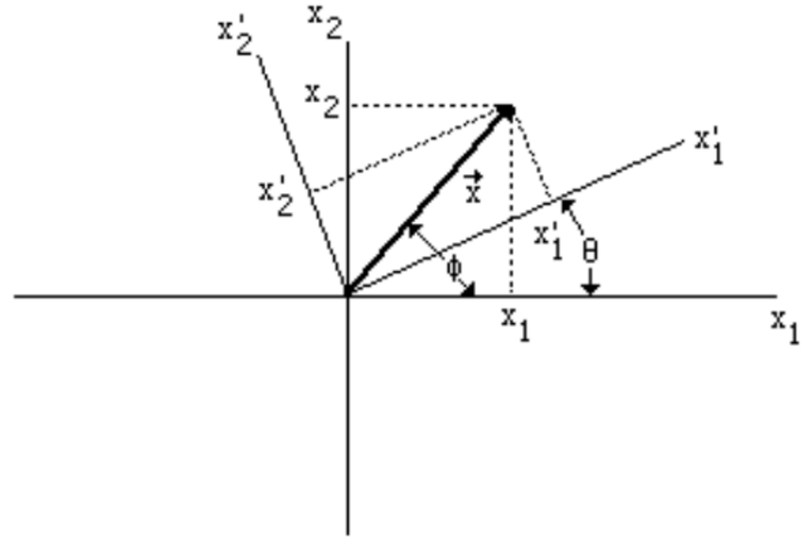
or

$$x'_1 = x_1 \cos\theta + x_2 \sin\theta$$

$$x'_2 = -x_1 \sin\theta + x_2 \cos\theta$$

$$x'_3 = x_3$$

which corresponds to the Figure below



Rotation about z-axis

If we let T_1 and T_2 be two such transformations, then $T_1 T_2$ is the transformation corresponding to T_2 followed by T_1.

A set of transformations forms a group when:

1. The product of two transformation in the group is also a transformation in the group.

2. The product is associative $T_3(T_2T_1) = (T_3T_2T_1)$.

3. An identity transformation T_0 exists such that $\vec{x} \to \vec{x}$ and $t \to t$ or $T_0T = T = TT_0$ for all transformations T.

4. An inverse transformation T^{-1} exists for every transformation such that $T^{-1}T = T_0 = TT^{-1}$.

A subset of transformations $T(\tau)$ depending on a real parameter τ is called a one-parameter subgroup if T_0 = the identity and $T(\tau_1 + \tau_2) = T(\tau_1)T(\tau_2)$.

Rotations about a fixed axis, as we saw above, form a one-parameter subgroup where the parameter is the angle of rotation.

Since products of the transformations from the three one-parameter subgroups, which correspond to rotations about the 1−, 2−, and 3−axes separately, include all possible rotations, there are three independent parameters(three angles) describing all rotations.

The total group of transformations has 10 parameters

1. rotations = 3 (3 angles)
2. space translations = 3 (3 components of $\vec{a}$)
3. Galilean boosts = 3 (3 components of $\vec{v}$)
4. time translation = 1 (parameter s)

Each of these ten parameters corresponds to a one-parameter subgroup.

A ten-parameter group transformation is then defined by the product of ten one-parameter subgroup transformations.

This means that in our discussions we need only consider the properties of one-parameter subgroup transformations in order to understand general group transformations.

In our earlier discussion we have already covered part of this topic when we derived the time evolution operator $\hat{U}\ (t)$, which clearly corresponds to the time translation transformation.

Our earlier results tell us that for each of the ten transformations there exists a linear, unitary operator $\hat{U}(\tau)$ on the state vectors and observables such that

$$|\psi\rangle \rightarrow |\psi'\rangle = \hat{U}(\tau)\,|\psi\rangle \ \text{ and } \hat{Q} \rightarrow \hat{Q}' = \hat{U}(\tau)\hat{Q}\hat{U}^{-1}(\tau)$$

where $\hat{U}(\tau)$ takes the general form

$$\hat{U}(\tau) = e^{i\tau\hat{G}}$$

and $\hat{G}$ = a Hermitian operator = the **generator of the transformation**.

The time evolution operator we derived earlier is a good example:

$$\text{if } |\psi(t)\rangle = \text{state vector at time } t, \text{ then for } t' = t + s$$

$$|\psi(t')\rangle = e^{i\tau \hat{H}} |\psi(t)\rangle$$

Generators of the Group Transformations

In general we can write

$$\hat{U}(s_\mu) = \prod_{\mu=1}^{10} e^{is_\mu \hat{K}_\mu}$$

where the different s_μ represent the ten parameters defining the group transformation represented by the operator $\hat{U}(s_\mu)$ and

$$\hat{K}_\mu = \hat{K}_\mu^\dagger = \text{ the Hermitian generators (there are 10)}$$

Now, let all the parameters become infinitesimally small so that we get infinitesimal unitary operators of the form (expanding the exponentials)

$$\hat{U}(s_\mu) = \prod_{\mu=1}^{10} e^{is_\mu \hat{K}_\mu} = \prod_{\mu=1}^{10} (\hat{I} + is_\mu \hat{K}_\mu)$$

or to first order in the parameters s_μ ,

$$\hat{U} = \hat{I} + i \sum_{\mu=1}^{10} s_\mu \hat{K}_\mu$$

Note that the inverse transformation corresponding to $e^{is_\mu \hat{K}_\mu}$ is $e^{-is_\mu \hat{K}_\mu}$.

We now construct the product transformation consisting of two infinitesimal transformations followed by their inverses to get

$$e^{i\epsilon \hat{K}_\mu} e^{i\epsilon \hat{K}_\nu} e^{-i\epsilon \hat{K}_\mu} e^{-i\epsilon \hat{K}_\nu} = \hat{I} + \epsilon^2 [\hat{K}_\nu, \hat{K}_\mu] + O(\epsilon^3)$$

where $[\hat{K}_\nu, \hat{K}_\mu] = \hat{K}_\nu \hat{K}_\mu - \hat{K}_\mu \hat{K}_\nu =$ commutator

The algebraic steps involved are shown below.

$$e^{i\epsilon\hat{K}_\mu} e^{i\epsilon\hat{K}_\nu} e^{-i\epsilon\hat{K}_\mu} e^{-i\epsilon\hat{K}_\nu}$$

$$= \left(\hat{I} + i\epsilon\hat{K}_\mu - \frac{1}{2}\epsilon^2\hat{K}_\mu^2\right)\left(\hat{I} + i\epsilon\hat{K}_\nu - \frac{1}{2}\epsilon^2\hat{K}_\nu^2\right) \times$$

$$\left(\hat{I} - i\epsilon\hat{K}_\mu - \frac{1}{2}\epsilon^2\hat{K}_\mu^2\right)\left(\hat{I} - i\epsilon\hat{K}_\nu - \frac{1}{2}\epsilon^2\hat{K}_\nu^2\right)$$

$$= \left(\hat{I} + i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2\hat{K}_\mu\hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)\right) \times$$

$$\left(\hat{I} - i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2\hat{K}_\mu\hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)\right)$$

$$= (\hat{I} + i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2\hat{K}_\mu\hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)$$

$$-i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2\hat{K}_\mu\hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)$$

$$+(i\epsilon(\hat{K}_\mu + \hat{K}_\nu))(-i\epsilon(\hat{K}_\mu + \hat{K}_\nu))$$

$$= \hat{I} - \epsilon^2\hat{K}_\mu^2 - \epsilon^2\hat{K}_\nu^2 - 2\epsilon^2\hat{K}_\mu\hat{K}_\nu + \epsilon^2\hat{K}_\mu^2 + \epsilon^2\hat{K}_\nu^2 + \epsilon^2\hat{K}_\mu\hat{K}_\nu + \epsilon^2\hat{K}_\nu\hat{K}_\mu$$

$$= \hat{I} + \epsilon^2\hat{K}_\nu\hat{K}_\mu - \epsilon^2\hat{K}_\mu\hat{K}_\nu$$

$$= \hat{I} + \epsilon^2[\hat{K}_\nu, \hat{K}_\mu] + O(\epsilon^3)$$

We have expanded to 2nd order in ε in this derivation to explicitly see all the higher order terms cancel out. In general, this is not necessary.

But remember that these transformations are part of a group and therefore the product of the four transformations must also be a transformation $\hat{W}$ in the same group.

Actually, it only needs to be a member of the group to within an arbitrary phase factor of the form $e^{i\alpha}$ (remember that the unitary transformations only need to preserve $|\langle\phi|\psi\rangle|$).

We then have

$$e^{i\alpha}\hat{W} = \hat{I} + \epsilon^2[\hat{K}_\nu, \hat{K}_\mu]$$

$$(1 + i\alpha)\left(\hat{I} + i\sum_\mu s_\mu \hat{K}_\mu\right) = \hat{I} + \epsilon^2[\hat{K}_\nu, \hat{K}_\mu]$$

$$i\alpha\hat{I} + i\sum_\mu s_\mu \hat{K}_\mu = \epsilon^2[\hat{K}_\nu, \hat{K}_\mu]$$

where we have used

$$\hat{W} = e^{i\sum_\mu s_\mu \hat{K}_\mu}$$

and expanded all the exponentials on the left-hand side of the equation to first order.

Therefore, the most general mathematical statement that we can make based on this result is that the commutator must take the form

$$[\hat{K}_\nu, \hat{K}_\mu] = i\sum_\lambda c_{\mu\nu}^\lambda \hat{K}_\mu + ib_{\mu\nu}\hat{I}$$

where the real numbers $c_{\mu\nu}^\lambda$ = the structure constants (of the group) and the term involving the identity operator just corresponds to the existence of an arbitrary phase factor.

The structure factors and the *b*'s are completely determined by the group algebra as we shall see.

By convention, we define the transformations as follows:

1. *Rotation* about the α-axis $(\alpha = 1, 2, 3)$

$$\vec{x} \to \mathbb{R}_\alpha(\theta_\alpha)\vec{x}$$

corresponds to the group operator

$$\hat{U} = e^{-i\theta_\alpha \hat{J}_\alpha}$$

where $\hat{J}_\alpha$ = generators $(\alpha = 1, 2, 3)$

2. *Displacement* along the α-axis $(\alpha = 1, 2, 3)$

$$x_\alpha \to x_\alpha + a_\alpha$$

corresponds to the group operator

$$\hat{U} = e^{-ia_\alpha \hat{P}_\alpha}$$

where $\hat{P}_\alpha$ = generators $(\alpha = 1, 2, 3)$

3. *Velocity boost* along the α-axis $(\alpha = 1, 2, 3)$

$$x_\alpha \to x_\alpha + v_\alpha t$$

corresponds to the group operator

$$\hat{U} = e^{iv_\alpha \hat{G}_\alpha}$$

where $\hat{G}_\alpha$ = generators $(\alpha = 1, 2, 3)$

4. *Time displacement*

$$t \to t + s$$

corresponds to the group operator

$$\hat{U} = e^{is\hat{H}}$$

where $\hat{H}$ = generator = Hamiltonian

Commutators and Identities

Initially, we will ignore the extra $\hat{I}$ term in the next equation below

$$[\hat{K}_\nu, \hat{K}_\mu] = i\sum_\lambda c^\lambda_{\mu\nu}\hat{K}_\mu + ib_{\mu\nu}\hat{I}$$

in our discussions and then include its effect(if any) later on.

We can determine some of the commutators using physical arguments as follows:

1. space displacements along different axes are independent of each other which implies that

$$[\hat{P}_\alpha, \hat{P}_\beta] = 0$$

2. space displacements are independent of time displacements which implies that

$$[\hat{P}_\alpha, \hat{H}] = 0$$

3. velocity boosts along different axes are independent of each other which implies that

$$[\hat{G}_\alpha, \hat{G}_\beta] = 0$$

4. rotations are independent of time displacements which implies that

$$[\hat{J}_\alpha, \hat{H}] = 0$$

5. space displacements and velocity boosts along a given axis are independent of rotations about that axis

$$[\hat{J}_\alpha, \hat{P}_\alpha] = 0 = [\hat{J}_\alpha, \hat{G}_\alpha]$$

6. obviously

$$[\hat{H}, \hat{H}] = 0$$

This leaves us to consider these remaining unknown commutators:

$$[\hat{J}_\alpha, \hat{P}_\beta], [\hat{G}_\alpha, \hat{H}], [\hat{J}_\alpha, \hat{J}_\beta], [\hat{G}_\alpha, \hat{J}_\beta], [\hat{G}_\alpha, \hat{P}_\beta]$$

Let us consider $[\hat{G}_1, \hat{H}]$ first.

The general procedure is as follows.

We write down a product of four operators consisting of the product of two operators representing a velocity boost in the 1-direction and a time translation and their inverses (as we did earlier).

Now these four successive transformation correspond to these changes of the coordinates:

$$\begin{aligned}
(x_1, x_2, x_3, t) &\to (x_1 - \epsilon t, x_2, x_3, t) && \text{Lorentz boost at time t} \\
&\to (x_1 - \epsilon t, x_2, x_3, t - \epsilon) && \text{Time translation - only affects t} \\
&\to (x_1 - \epsilon t + \epsilon(t - \epsilon), x_2, x_3, t - \epsilon) && \text{Lorentz boost at time t-}\epsilon \\
&\to (x_1 - \epsilon^2, x_2, x_3, t) && \text{Time translation - only affects t}
\end{aligned}$$

This last result just corresponds to a space displacement $-\varepsilon^2$ along the 1-axis, so equating the product of four transformations to a space translation, we have the result

$$\begin{aligned}
& e^{i\epsilon\hat{H}} e^{i\epsilon\hat{G}_1} e^{-i\epsilon\hat{H}} e^{-i\epsilon\hat{G}_1} \\
&\quad = (\hat{I} + i\epsilon\hat{H})(\hat{I} + i\epsilon\hat{G}_1)(\hat{I} - i\epsilon\hat{H})(\hat{I} - i\epsilon\hat{G}_1) \\
&\quad = \hat{I} + \epsilon^2[\hat{G}_1, \hat{H}] + O(\epsilon^3) \\
&\quad = e^{-i(-\epsilon^2)\hat{P}_1} = \hat{I} + i\epsilon^2\hat{P}_1
\end{aligned}$$

so that we find the result for the commutator

$$[\hat{G}_1, \hat{H}] = i\hat{P}_1$$

In general, we get (using this same procedure)

$$[\hat{G}_\alpha, \hat{H}] = i\hat{P}_\alpha$$

So we have determined one of the unknown commutators.

Now let us determine [$\hat{J}_\alpha, \hat{J}_\beta$] using the same type of procedure.

For a rotation we saw that

$$x_j \to \sum_k R_{jk} x_k$$

where

$$\mathbb{R}_1(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}$$

$$\mathbb{R}_2(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$$

$$\mathbb{R}_3(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

For small θ we have

$$\mathbb{R}_\alpha(\theta) = \hat{I} - i\theta M_\alpha$$

where M_α is determined by expanding the exponential to 1st order as

$$M_\alpha = i \left. \frac{d\mathbb{R}_\alpha}{d\theta} \right|_{\theta=0}$$

$$M_1(\theta) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & -i & 0 \end{pmatrix}$$

$$M_2(\theta) = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}$$

$$M_3(\theta) = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Then we have

$$\begin{aligned} &\mathbb{R}_2(-\epsilon)\mathbb{R}_1(-\epsilon)\mathbb{R}_2(\epsilon)\mathbb{R}_1(\epsilon) \\ &\quad = \hat{I} + \epsilon^2[M_1, M_2] = \hat{I} + i\epsilon^2 M_3 \\ &\quad = \mathbb{R}_3(-\epsilon^2) \end{aligned}$$

or the product of four rotations is equivalent to a single rotation, which implies that

$$[\hat{J}_1, \hat{J}_2] = i\hat{J}_3$$

and in general

$$[\hat{J}_\alpha, \hat{J}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{J}_\gamma$$

where we are using the Einstein summation convention for repeated indices and $\varepsilon_{\alpha\beta\gamma}$ is the antisymmetric permutation symbol we introduced earlier with the properties

$$\epsilon_{\alpha\beta\gamma} = \begin{cases} 1 & \text{if } \alpha\beta\gamma \text{ is an even permutation of } 123 \\ -1 & \text{if } \alpha\beta\gamma \text{ is an odd permutation of } 123 \\ 0 & \text{if any two indices are the same} \end{cases}$$

Finally, we consider

$$e^{i\epsilon\hat{G}_2} e^{i\epsilon\hat{J}_1} e^{-i\epsilon\hat{G}_2} e^{-i\epsilon\hat{J}_1}$$

$$= (\hat{I} + i\epsilon G_2)(\hat{I} + i\epsilon\hat{G}_1)(\hat{I} - i\epsilon G_2)(\hat{I} - i\epsilon\hat{G}_1)$$

$$= \hat{I} + \epsilon^2[\hat{J}_1, \hat{G}_2] + O(\epsilon^3)$$

This involves a rotation by ε about the 1-axis and a velocity boost of ε along the 2-axis.

This product transformation changes the coordinates as follows:

$$(x_1, x_2, x_3) \to (x_1, x_2\cos\epsilon + x_3\sin\epsilon, -x_2\sin\epsilon + x_3\cos\epsilon)$$

$$\to (x_1, x_2\cos\epsilon + x_3\sin\epsilon - \epsilon t, -x_2\sin\epsilon + x_3\cos\epsilon)$$

$$\to (x_1, x_2 + \epsilon t\cos\epsilon, x_3 + \epsilon t\sin\epsilon)$$

$$\to (x_1, x_2, x_3 + \epsilon^2 t) \text{ to } 2^{nd} \text{ order in } \epsilon$$

This is the same as

$$e^{i\epsilon^2\hat{G}_3} = \hat{I} + i\epsilon^2\hat{G}_3$$

Thus, we have

$$[\hat{J}_1, \hat{G}_2] = i\hat{G}_3$$

or in general,

$$[\hat{J}_\alpha, \hat{G}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{G}_\gamma$$

In a similar way,

$$[\hat{J}_\alpha, \hat{P}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{P}_\gamma$$

Summarizing so far we have

$$[\hat{P}_\alpha, \hat{P}_\beta] = 0 \ , \ [\hat{P}_\alpha, \hat{H}] = 0 \ , \ [\hat{G}_\alpha, \hat{G}_\beta] = 0 \ , \ [\hat{J}_\alpha, \hat{H}] = 0$$

$$[\hat{J}_\alpha, \hat{P}_\alpha] = 0 \ , \ [\hat{J}_\alpha, \hat{G}_\alpha] = 0 \ , \ [\hat{H}, \hat{H}] = 0$$

$$[\hat{G}_\alpha, \hat{H}] = i\hat{P}_\alpha \ , \ [\hat{J}_\alpha, \hat{J}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{J}_\gamma$$

$$[\hat{J}_\alpha, \hat{G}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{G}_\gamma \ , \ [\hat{J}_\alpha, \hat{P}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{P}_\gamma$$

Before figuring out the last unknown commutator $[\hat{G}_\alpha, \hat{P}_\beta]$, we need to see whether the additional $\hat{I}$ term has an effect on any of the commutators we have already determined.

In the book I show that it does not.

The proofs use a very useful identity among commutators call the Jacobi Identity:

$$[[\hat{A}, \hat{B}], \hat{C}] = [[\hat{C}, \hat{B}], \hat{A}] + [[\hat{A}, \hat{C}], \hat{B}]$$

Finally, we are left to consider the unknown commutator $[\hat{G}_\alpha, \hat{P}_\beta]$.

Now using the fact that $[c\hat{I}, \hat{P}_1] = 0$ we have

$$\begin{aligned} i[\hat{G}_3, \hat{P}_1] &= [[\hat{J}_1, \hat{G}_2], \hat{P}_1] + [c\hat{I}, \hat{P}_1] = [[\hat{J}_1, \hat{G}_2], \hat{P}_1] \\ &= [[\hat{P}_1, \hat{G}_2], \hat{J}_1] + [[\hat{J}_1, \hat{P}_1], \hat{G}_2] = [[\hat{P}_1, \hat{G}_2], \hat{J}_1] \end{aligned}$$

which has the solution

$$[\hat{G}_\alpha, \hat{P}_\beta] = 0 \quad \alpha \neq \beta$$

In addition, we have

$$
\begin{aligned}
i[\hat{G}_3, \hat{P}_3] &= [[\hat{J}_1, \hat{G}_2], \hat{P}_3] + [c\hat{I}, \hat{P}_3] = [[\hat{J}_1, \hat{G}_2], \hat{P}_3] \\
&= [[\hat{P}_3, \hat{G}_2], \hat{J}_1] + [[\hat{J}_1, \hat{P}_3], \hat{G}_2] \\
&= -i[\hat{P}_2, \hat{G}_2]
\end{aligned}
$$

or

$$[\hat{G}_3, \hat{P}_3] = [\hat{G}_2, \hat{P}_2]$$

and, in general

$$[\hat{G}_\alpha, \hat{P}_\alpha] = [\hat{G}_\beta, \hat{P}_\beta]$$

The only way to satisfy all of these commutators is to have the result

$$[G_\alpha, P_\beta] = \delta_{\alpha\beta} M I$$

The value of M is undetermined.

It cannot be eliminated by including multiples of $\hat{I}$ in any of the other commutators.

It must have some real physical significance and we will identify it shortly.

Identification of Operators with Observables

We now use the dynamics of a free particle, which is a physical system that is invariant under the Galilei group of space-time transformations, to identify the operators representing the dynamical variables or observables in that case.

We assume that there exists a position operator (as we discussed earlier)

$$\hat{\mathbf{Q}} = (\hat{Q}_1, \hat{Q}_2, \hat{Q}_3)$$

(**boldface** = multi-component or vector operator) where

$$\hat{Q}_\alpha |\vec{x}\rangle = x_\alpha |\vec{x}\rangle \quad , \quad \alpha = 1, 2, 3$$

The position operator has an unbounded ($-\infty < x_\alpha < \infty$), continuous spectrum.

The three operators $\hat{Q}_\alpha$ have a common set of eigenvectors and, thus, they commute

$$[\hat{Q}_\alpha, \hat{Q}_\beta] = 0$$

We now **assume** that there also exists a velocity operator

$$\hat{\mathbf{V}} = (\hat{V}_1, \hat{V}_2, \hat{V}_3)$$

such that we can make the following statement about expectation values (*this is an assumption*)

$$\frac{d}{dt}\langle \hat{\mathbf{Q}} \rangle = \langle \hat{\mathbf{V}} \rangle$$

for any state vector.

It is important to note that each treatment of quantum mechanics must make some assumption at this point.

Although, they might outwardly look like different assumptions, they clearly must be equivalent since they result in the same theory with the same predictions.

For a pure state $|\psi(t)\rangle$ we would then have

$$\langle\psi(t)|\,\hat{\mathbf{V}}\,|\psi(t)\rangle = \frac{d}{dt}\left(\langle\psi(t)|\,\hat{\mathbf{Q}}\,|\psi(t)\rangle\right)$$
$$= \left[\frac{d}{dt}\langle\psi(t)|\right]\hat{\mathbf{Q}}\,|\psi(t)\rangle + \langle\psi(t)|\,\hat{\mathbf{Q}}\left[\frac{d}{dt}|\psi(t)\rangle\right]$$

Now a time displacement corresponds to $t \rightarrow t' = t + s$.

To use this transformation, we first have to figure out a rule for the transformation of the ket vector argument.

When we represent the abstract vector $|\psi\rangle$ as a function of spacetime we must be very careful about defining its properties.

We derived this result earlier(I skipped it earlier), but it is so important that we do it again here

Given $|\psi\rangle$, we have that

$$\psi(\vec{x},t) = \langle\vec{x},t\,|\,\psi\rangle$$

Operating on the abstract vector with the time translation operator $e^{is\hat{H}}|\psi\rangle$ then gives

$$\langle\vec{x},t|\,e^{is\hat{H}}\,|\psi\rangle$$

Now

$$e^{-is\hat{H}}\,|\vec{x},t\rangle = |\vec{x},t-s\rangle$$

and therefore we get

$$\langle\vec{x},t|\,e^{is\hat{H}}\,|\psi\rangle = \langle\vec{x},t-s\,|\,\psi\rangle = \psi(\vec{x},t-s)$$

or

$$|\psi(t)\rangle \rightarrow e^{is\hat{H}}\,|\psi(t)\rangle = |\psi(t-s)\rangle$$

Here we have taken the so-called **active point of view** that the state is translated relative to a fixed coordinate system.

Now, we let $s = t$ to get

$$|\psi(t)\rangle \rightarrow e^{it\hat{H}}\,|\psi(t)\rangle = |\psi(0)\rangle$$

or

$$|\psi(t)\rangle = e^{-it\hat{H}}\,|\psi(0)\rangle$$

and

$$\frac{d}{dt}\,|\psi(t)\rangle = -i\hat{H}\,|\psi(t)\rangle$$

as we had already found in an earlier discussion.

Using this result in the above equation(assumed) we then find

$$\begin{aligned}\langle\psi(t)|\,\hat{\mathbf{V}}\,|\psi(t)\rangle &= i\,\langle\psi(t)|\,\hat{H}\hat{\mathbf{Q}}\,|\psi(t)\rangle - i\,\langle\psi(t)|\,\hat{\mathbf{Q}}\hat{H}\,|\psi(t)\rangle \\ &= i\,\langle\psi(t)|\,[\hat{H},\hat{\mathbf{Q}}]\,|\psi(t)\rangle\end{aligned}$$

which says that

$$\hat{\mathbf{V}} = i[\hat{H},\hat{\mathbf{Q}}]$$

is a **valid** velocity operator for a free particle.

However, since we **do not know** $\hat{H}$, **we are still stuck at the starting line**.

Now a space displacement $\vec{x} \rightarrow \vec{x}' = \vec{x} + \vec{a}$ corresponds to

$$|\vec{x}\rangle \rightarrow |\vec{x}'\rangle = e^{-i\sum_\alpha a_\alpha \hat{P}_\alpha} |\vec{x}\rangle = |\vec{x} + \vec{a}\rangle$$

Again, in this active point of view, the state is displaced relative to coordinate system.

We then have, from our earlier discussion

$$\hat{\mathbf{Q}} \rightarrow \hat{\mathbf{Q}}' = e^{-i\sum_\alpha a_\alpha \hat{P}_\alpha} \hat{\mathbf{Q}} e^{i\sum_\alpha a_\alpha \hat{P}_\alpha}$$

where

$$\hat{Q}'_\alpha |\vec{x}\rangle' = x_\alpha |\vec{x}\rangle' \text{ or } \hat{Q}'_\alpha |\vec{x} + \vec{a}\rangle = x_\alpha |\vec{x} + \vec{a}\rangle$$

Since

$$\hat{Q}_\alpha |\vec{x}\rangle = x_\alpha |\vec{x}\rangle \text{ or } \hat{Q}_\alpha |\vec{x} + \vec{a}\rangle = (x_\alpha + a_\alpha) |\vec{x} + \vec{a}\rangle$$

we must have

$$(\hat{Q}_\alpha - a_\alpha \hat{I}) |\vec{x} + \vec{a}\rangle = x_\alpha |\vec{x} + \vec{a}\rangle$$

which says that

$$\hat{Q}'_\alpha = \hat{Q}_\alpha - a_\alpha \hat{I} \text{ or } \hat{\mathbf{Q}}' = \hat{\mathbf{Q}} - \vec{a}\hat{I}$$

Now we need to work out how to use an operator of the form

$$e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}}$$

There are two ways to do this.

The first way recognizes that it is the solution of the 1st-order linear differential equation

$$i\frac{d}{du} e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}} = -e^{iu\hat{A}} \hat{A} \hat{B} e^{-iu\hat{A}} + e^{iu\hat{A}} \hat{B} \hat{A} e^{-iu\hat{A}}$$

$$= e^{iu\hat{A}} [\hat{A}, \hat{B}] e^{-iu\hat{A}} = [e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}}, \hat{A}]$$

with boundary condition

$$e^{iu\hat{A}}\hat{B}e^{-iu\hat{A}} = \hat{B} \text{ at } u = 0$$

A second way is to expand the exponentials in power series and regroup the terms.

In both cases, we get

$$e^{iu\hat{A}}\hat{B}e^{-iu\hat{A}} = \hat{B} - iu[\hat{B},\hat{A}] - \frac{u^2}{2!}[[\hat{B},\hat{A}],\hat{A}] - \dots$$

Using this last relation we can find useful results like

$$\begin{aligned} e^{iu\hat{J}_1}\hat{J}_3 e^{-iu\hat{J}_1} &= \hat{J}_3 - iu[\hat{J}_3,\hat{J}_1] - \frac{u^2}{2!}[[\hat{J}_3,\hat{J}_1],\hat{J}_1] - \dots \\ &= \hat{J}_3 - iu(i\hat{J}_2) - \frac{u^2}{2!}[i\hat{J}_2,\hat{J}_1] - \dots \\ &= \hat{J}_3 + u\hat{J}_2 - \frac{u^2}{2!}\hat{J}_3 - \frac{u^3}{3!}\hat{J}_2 + \frac{u^4}{4!}\hat{J}_3 - \dots \\ &= \hat{J}_3\left(1 - \frac{u^2}{2!} + \frac{u^4}{4!} - \dots\right) + \hat{J}_2\left(u - \frac{u^3}{3!} + \frac{u^5}{5!} - \dots\right) \\ &= \hat{J}_3\cos u + \hat{J}_2\sin u \end{aligned}$$

Now we finally arrive at the result we are looking for

$$\begin{aligned} \hat{Q}'_\alpha = \hat{Q}_\alpha - a_\alpha\hat{I} &= e^{-i\sum_\beta a_\beta\hat{P}_\beta}\hat{Q}_\alpha e^{i\sum_\beta a_\beta\hat{P}_\beta} \\ &= \hat{Q}_\alpha + i\sum_\beta a_\beta[\hat{Q}_\alpha,\hat{P}_\beta] + \dots \end{aligned}$$

which implies that

$$[\hat{Q}_\alpha,\hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I}$$

This solution, not only satisfies the equation to 1st-order as above, but implies that all higher order terms are explicitly equal to zero.

Thus, the solution is exact.

This equation is one of the most important results in the theory of quantum mechanics.

Now we continue the derivation of that last elusive commutator.

A rotation through an infinitesimal angle θ about an axis along the unit vector $\hat{\mathbf{n}}$ has the effect

$$\vec{x} \to \vec{x}' = \vec{x} + \theta \hat{\mathbf{n}} \times \vec{x}$$

The corresponding transformation of the position eigenvectors is

$$|\vec{x}\rangle \to |\vec{x}'\rangle = e^{-i\theta \hat{\mathbf{n}} \cdot \hat{\mathbf{J}}} |\vec{x}\rangle$$

and for the position operator

$$\begin{aligned} \hat{Q}_\alpha \to \hat{Q}'_\alpha &= e^{-i\theta \hat{\mathbf{n}} \cdot \hat{\mathbf{J}}} \hat{Q}_\alpha e^{i\theta \hat{\mathbf{n}} \cdot \hat{\mathbf{J}}} \\ &= \hat{Q}_\alpha - i\theta [\hat{\mathbf{n}} \cdot \hat{\mathbf{J}}, \hat{Q}_\alpha] + O(\theta^2) \end{aligned}$$

We note that $|\vec{x}'\rangle$ = an eigenvector of $\hat{Q}'$ and $|\vec{x}\rangle$ = an eigenvector of $\hat{Q}$ (different eigenvalues), but they are the **same** vector! (think about rotations of the axes).

Now as from earlier,

$$\hat{Q}'_\alpha |\vec{x}\rangle' = x_\alpha |\vec{x}\rangle'$$

$$\begin{aligned} \hat{Q}_\alpha |\vec{x}'\rangle = x'_\alpha |\vec{x}'\rangle &= (\vec{x} + \theta \hat{\mathbf{n}} \times \vec{x})_\alpha |\vec{x}'\rangle \\ &= (\hat{\mathbf{Q}}' + \theta \hat{\mathbf{n}} \times \hat{\mathbf{Q}}')_\alpha |\vec{x}\rangle' \end{aligned}$$

Now since the vectors $|\vec{x}'\rangle = |\vec{x}\rangle'$ are a complete set, we must have

$$\hat{\mathbf{Q}} = \hat{\mathbf{Q}}' + \theta\hat{\mathbf{n}} \times \hat{\mathbf{Q}}'$$

or inverting this expression to 1st-order we have

$$\hat{\mathbf{Q}}' = \hat{\mathbf{Q}} - \theta\hat{\mathbf{n}} \times \hat{\mathbf{Q}}$$

Therefore, we find

$$[\hat{\mathbf{n}} \cdot \hat{\mathbf{J}}, \hat{Q}_\alpha] = -i\hat{\mathbf{n}} \times \hat{\mathbf{Q}}$$

For an arbitrary unit vector $\hat{\mathbf{u}}$ this says that

$$[\hat{\mathbf{n}} \cdot \hat{\mathbf{J}}, \hat{\mathbf{u}} \cdot \hat{Q}_\alpha] = -i\hat{\mathbf{u}} \cdot (\hat{\mathbf{n}} \times \hat{\mathbf{Q}}) = i(\hat{\mathbf{n}} \times \hat{\mathbf{u}}) \cdot \hat{\mathbf{Q}}$$

or

$$[\hat{J}_\alpha, \hat{Q}_\beta] = i(\hat{e}_\alpha \times \hat{e}_\beta) \cdot \hat{\mathbf{Q}} = i(\epsilon_{\alpha\beta\gamma}\hat{e}_\gamma) \cdot \hat{\mathbf{Q}} = i\epsilon_{\alpha\beta\gamma}\hat{Q}_\gamma$$

We note that this result is not only true for the components of the position operator, but it is true for the components of **any** vector operator $\hat{\mathbf{A}}$.

$$[\hat{J}_\alpha, \hat{A}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{A}_\gamma$$

In a similar manner, since $\hat{\mathbf{G}}$ generates a displacement in velocity space and we have

$$\hat{\mathbf{V}}' = \hat{\mathbf{V}} - \vec{v}\hat{I} = e^{i\vec{v}\cdot\hat{\mathbf{G}}}\hat{\mathbf{V}}e^{-i\vec{v}\cdot\hat{\mathbf{G}}}$$

This is the same way the $\hat{\mathbf{Q}}$ and $\hat{\mathbf{P}}$ operators behaved earlier.

Now the unitary operator $\hat{U}(\vec{v}) = e^{i\vec{v}\cdot\hat{\mathbf{G}}}$ describes the instantaneous ($t = 0$) effect of a transformation to a frame of reference moving at velocity $\vec{\mathrm{v}}$ with respect to the original frame.

It affects the $\hat{\mathbf{V}}$ operator as in the last commutator.

$$[\hat{J}_\alpha, \hat{V}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{V}_\gamma$$

Due to its instantaneous nature we must also have

$$\hat{U}\hat{\mathbf{Q}}\hat{U}^{-1} = \hat{\mathbf{Q}} \text{ or } [\hat{G}_\alpha, \hat{Q}_\beta] = 0$$

Now, $\hat{Q}$, the position operator, is clearly identified with an observable of the physical state.

After much work we have determined the commutators of $\hat{Q}$ with all the symmetry generators of the Galilei group.

We now have $[\hat{G}_\alpha, \hat{Q}_\beta] = 0 \quad , \quad [\hat{G}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}M\hat{I} \quad , \quad [\hat{Q}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I}$

A **possible solution** of these equations is

$$\hat{G}_\alpha = M\hat{Q}_\alpha$$

We cannot show this is a unique solution.

We note, however, at this point that it is certainly true (using the earlier commutators) that

$$[\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{P}}] = 0 \text{ and } [\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{Q}}] = 0$$

We must now turn to special cases in order to learn more about the physical meaning of the generators.

Before discussing the special cases, we must again digress to add some mathematics (discussed earlier in the book) and also cover some new mathematical ideas that we will need.

A subspace M *reduces* a linear operator $\hat{A}$ if $\hat{A}|\psi\rangle$ is in M for every $|\psi\rangle$ in M and $\hat{A}|\phi\rangle$ is in $M^{\perp}$ for every $|\phi\rangle$ in $M^{\perp}$.

A set of operators is *reducible* if there is a subspace, other than the whole space or the subspace containing only the null vector, which *reduces* every operator in the set. Otherwise, we say that the set is *irreducible.*

A subspace M is *invariant* under a set of operators if $\hat{A}|\psi\rangle$ is in M for every operator in the set and every vector in M.

Thus, a subspace M reduces a set of operators *if and only if* M and $M^{\perp}$ are invariant under the set of operators.

A set of operators is *symmetric* if $\hat{A}^{\dagger}$ is in the set for every operator $\hat{A}$ in the set.

If a subspace is invariant under a symmetric set of operators, then it reduces the set of operators.

Schur's Lemma

A symmetric set of bounded or Hermitian operators is irreducible if and only if multiples of $\hat{I}$ are the only bounded operators which commute with all operators in the set.

Example 0

The commutator $[\hat{Q}_\alpha, \hat{Q}_\beta] = 0$ says that the set of operators $\{\hat{Q}_\alpha\ , \alpha = 1,2,3\}$ is a complete set of commuting operators.

Since $[\hat{Q}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I}$ any function of $\hat{Q}_\alpha$ that commutes with the $\hat{P}_\alpha$ must be a multiple of $\hat{I}$.

Therefore, the set $\{\hat{Q}_1, \hat{Q}_2, \hat{Q}_3, \hat{P}_1, \hat{P}_2, \hat{P}_3\}$ is irreducible.

In other words, if an operator commutes with the $\hat{Q}_\alpha$, then it is not a function of $\hat{P}_\alpha$ since $[\hat{Q}_\alpha, \hat{P}_\beta] \neq 0$.

If an operator commutes with the $\hat{P}_\alpha$, then it is not a function of $\hat{Q}_\alpha$, for the same reason.

If an operator is independent of the set $\{\hat{Q}_\alpha, \hat{P}_\beta\}$ and there are no internal(not dependent on $\{\hat{Q}_\alpha, \hat{P}_\beta\}$) degrees of freedom, then it must be a multiple of $\hat{I}$.

Example 1 : Free Particle - no internal degrees of freedom

Now, as we stated earlier that

$$[\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{P}}] = 0 \text{ and } [\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{Q}}] = 0$$

and therefore

$$\hat{\mathbf{G}} - M\hat{\mathbf{Q}} = \text{ multiple of } \hat{I}$$

or

$$\hat{G}_\alpha = M\hat{Q}_\alpha + c_\alpha\hat{I}$$

But $\hat{G}_\alpha$ is a component of a vector operator and therefore must satisfy

$$[\hat{J}_\alpha, \hat{G}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{G}_\gamma$$

The first term $M\hat{G}_\alpha$ satisfies this relation, but the $c_\alpha \hat{I}$ term cannot unless we choose $c_\alpha = 0$.

Therefore, we do not have any extra multiples of $\hat{I}$ and we find that

$$\hat{G}_\alpha = M\hat{Q}_\alpha$$

when there are no internal degrees of freedom(a free particle in this case).

In a similar manner, we can show that

$$[\hat{\mathbf{J}} - \hat{\mathbf{Q}} \times \hat{\mathbf{P}}, \hat{\mathbf{P}}] = 0 \text{ and } [\hat{\mathbf{J}} - \hat{\mathbf{Q}} \times \hat{\mathbf{P}}, \hat{\mathbf{Q}}] = 0$$

which then implies that

$$\hat{\mathbf{J}} - \hat{\mathbf{Q}} \times \hat{\mathbf{P}} = \vec{c}\hat{I}$$

But since

$$[\hat{J}_\alpha, \hat{J}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{J}_\gamma$$

again we are forced to choose $\vec{c} = 0$ and thus

$$\hat{\mathbf{J}} = \hat{\mathbf{Q}} \times \hat{\mathbf{P}}$$

when there are no internal degrees of freedom.

The remaining generator we need to identify is $\hat{H}$.

It must satisfy

$$[\hat{G}_\alpha, \hat{H}] = i\hat{P}_\alpha \rightarrow [M\hat{Q}_\alpha, \hat{H}] = i\hat{P}_\alpha \rightarrow [\hat{Q}_\alpha, \hat{H}] = i\frac{\hat{P}_\alpha}{M}$$

A solution of this equation is given by

$$\hat{H} = \frac{\hat{\mathbf{P}} \cdot \hat{\mathbf{P}}}{2M} = \frac{\hat{\mathbf{P}}^2}{2M}$$

as can be seen below

$$\left[\hat{Q}_\alpha, \frac{1}{2M}\hat{\mathbf{P}}^2\right] = \left[\hat{Q}_\alpha, \frac{1}{2M}\sum_\beta \hat{P}_\beta^2\right] = \frac{1}{2M}\sum_\beta [\hat{Q}_\alpha, \hat{P}_\beta^2]$$

$$= \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta^2 - \hat{P}_\beta^2 \hat{Q}_\alpha) = \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta \hat{P}_\beta - \hat{P}_\beta \hat{P}_\beta \hat{Q}_\alpha)$$

$$= \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta \hat{P}_\beta - \hat{P}_\beta (\hat{Q}_\alpha \hat{P}_\beta - i\delta_{\alpha\beta}\hat{I}))$$

$$= \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta \hat{P}_\beta - \hat{P}_\beta \hat{Q}_\alpha \hat{P}_\beta + i\delta_{\alpha\beta}\hat{P}_\beta)$$

$$= \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta \hat{P}_\beta - (\hat{Q}_\alpha \hat{P}_\beta - i\delta_{\alpha\beta}\hat{I})\hat{P}_\beta + i\delta_{\alpha\beta}\hat{P}_\beta)$$

$$= \frac{1}{2M}\sum_\beta 2i\delta_{\alpha\beta}\hat{P}_\beta = \frac{i\hat{P}_\alpha}{M}$$

This result implies that $\hat{H} - \hat{\mathbf{P}}\cdot\hat{\mathbf{P}}/2M$ commutes with $\hat{\mathbf{Q}}$ and since $[\hat{H}, \hat{\mathbf{P}}] = 0$, it also commutes with $\hat{\mathbf{P}}$.

Therefore, it is a multiple of $\hat{I}$, and we find

$$\hat{H} - \frac{\hat{\mathbf{P}}\cdot\hat{\mathbf{P}}}{2M} = E_0\hat{I} \quad , \quad E_0 = \text{ constant}$$

or

$$\hat{H} = \frac{\hat{\mathbf{P}}\cdot\hat{\mathbf{P}}}{2M} + E_0\hat{I}$$

Now, earlier we found that

$$[\hat{Q}_\alpha, \hat{H}] = i\hat{V}_\alpha$$

which now implies that

$$[\hat{Q}_\alpha, \hat{H}] = \left[\hat{Q}_\alpha, \frac{1}{2M}\hat{\mathbf{P}}^2\right] = \frac{i\hat{P}_\alpha}{M} = i\hat{V}_\alpha$$

Thus,

$$\hat{V}_\alpha = \frac{\hat{P}_\alpha}{M} \text{ or } \hat{\mathbf{V}} = \frac{\hat{\mathbf{P}}}{M}$$

Summarizing, we have found that

$$\hat{\mathbf{P}} = M\hat{\mathbf{V}} \quad , \quad \hat{H} = \frac{M}{2}\hat{\mathbf{V}} \cdot \hat{\mathbf{V}} + E_0 \quad , \quad \hat{\mathbf{J}} = \hat{\mathbf{Q}} \times M\hat{\mathbf{V}}$$

where

$\hat{\mathbf{V}}$ = the velocity operator and $\hat{\mathbf{Q}}$ = the position operator

This implies, since we are talking about a **free particle**, that

M must be *proportional* to the mass of the free particle

i.e., suppose that

$$M = \text{constant} \times \text{mass} = \beta m$$

then we must have

$$\hat{P}_\alpha = \text{constant} \times \text{linear momentum} = \beta\hat{p}_\alpha$$

$$\hat{H} = \text{constant} \times \text{energy} = \beta\hat{E}$$

$$\hat{J}_\alpha = \text{constant} \times \text{angular momentum} = \beta\hat{j}_\alpha$$

and all the earlier relations

$$\hat{\mathbf{P}} = M\hat{\mathbf{V}} \quad , \quad \hat{H} = \frac{M}{2}\hat{\mathbf{V}} \cdot \hat{\mathbf{V}} + E_0 \quad , \quad \hat{\mathbf{J}} = \hat{\mathbf{Q}} \times M\hat{\mathbf{V}}$$

still hold.

Now also remember that

$$[\hat{Q}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I}$$

which then implies that

$$[\hat{Q}_\alpha, \hat{p}_\beta] = i\frac{1}{\beta}\delta_{\alpha\beta}\hat{I}$$

This says that 1/β must have the units Joule-Sec.

With hindsight as to later developments, we will now choose

$$\hbar = \frac{1}{\beta}$$

and thus we finally obtain

$$[\hat{Q}_\alpha, \hat{p}_\beta] = i\hbar\delta_{\alpha\beta}\hat{I}$$

At this point, we do not know the numerical value of $\hbar$ (any value will do at this stage of our development).

It can only be determined by experiment.

Later we shall see that it is

$$\hbar = \frac{h}{2\pi}$$

where h = Planck's constant = 6.62×10^{-34} Joule-Sec.

$\hbar$ must be determined from experiment since it sets the scale of all quantum phenomena and the scale factors in physical theories cannot be known a priori (even though Kant thought just the opposite was true).

Example 2 : Free Particle - with Spin

Internal degrees of freedom are, by definition, independent of the center of mass degrees of freedom(Example 1).

This means they are represented by operators that are independent of both $\hat{\mathbf{Q}}$ and $\hat{\mathbf{P}}$, or that they are represented by operators that commute with both $\hat{\mathbf{Q}}$ and $\hat{\mathbf{P}}$.

The set of operators $\{\hat{\mathbf{Q}}, \hat{\mathbf{P}}\}$ is not irreducible in this case since an operator that commutes with the set may still be a function of the operators corresponding to the internal degrees of freedom.

The spin $\hat{\mathbf{S}}$ is defined to be an internal contribution to the total angular momentum of the system.

Therefore, we must modify the operator representing angular momentum to be

$$\hat{\mathbf{J}} = \hat{\mathbf{Q}} \times \hat{\mathbf{P}} + \hat{\mathbf{S}}$$

with $[\hat{\mathbf{Q}}, \hat{\mathbf{S}}] = 0 = [\hat{\mathbf{P}}, \hat{\mathbf{S}}]$.

We will study spin in great detail later in these lectures.

For now, let us see what we can say just based on the fact that spin is an angular momentum.

This means that the $\hat{S}_\alpha$ must have the same commutators among themselves as the $\hat{J}_\alpha$.

$$[\hat{S}_\alpha, \hat{S}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{S}_\gamma$$

Earlier we found the equation

$$[\hat{G}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}M\hat{I}$$

is satisfied by $\hat{\mathbf{G}} = M\hat{\mathbf{Q}}$ and

$$[\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{P}}] = 0 \text{ and } [\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{Q}}] = 0$$

implied that $\hat{\mathbf{G}} - M\hat{\mathbf{Q}} = \vec{a}\hat{I}$ and we decided that we must have $\vec{\mathrm{a}} = 0$.

Now, however, we could also have terms involving $\hat{\mathbf{S}}$ of the form

$$\hat{\mathbf{G}} - M\hat{\mathbf{Q}} = c\hat{\mathbf{S}}$$

Now we found earlier that $[\hat{G}_\alpha, \hat{G}_\beta] = 0$, which implies that $c = 0$ and thus, we still have $\hat{\mathbf{G}} = M\hat{\mathbf{Q}}$.

The previous argument to get

$$\hat{H} = \frac{\hat{\mathbf{P}} \cdot \hat{\mathbf{P}}}{2M} + E_0\hat{I}$$

is identical to before except now E_0 can be a function of $\hat{\mathbf{S}}$.

Now we must have

$$[\hat{\mathbf{J}}, \hat{H}] = 0 \rightarrow [\hat{\mathbf{S}}, E_0] = 0$$

which implies that $E_0 = \mathrm{c}\ \hat{\mathbf{S}} \cdot \hat{\mathbf{S}}$ is the only possibility.

This has no effect on the $\hat{\mathbf{V}}$ operator given by

$$\hat{\mathbf{V}} = i[\hat{H}, \hat{\mathbf{Q}}]$$

since $[E_0\ ,\ \hat{\mathbf{Q}}\ ] = 0$.

Therefore, the relation

$$\hat{\mathbf{V}} = \frac{\hat{\mathbf{P}}}{M}$$

remains valid.

So everything is the same as in Example 1 except that E_0 now corresponds to an internal (spin dependent) contribution to the energy.

Example 3 - A Particle Interacting with External Fields

We will only consider a **spinless** particle.

Interactions change the time evolution operator and as a consequence, the probability distributions for observables.

We **assume** that the equation of motion for the state vector **retains its form**

$$\frac{d}{dt}|\psi(t)\rangle = -i\hat{H}|\psi(t)\rangle$$

but the generator $\hat{H}$ changes to include the interactions.

We single out $\hat{H}$ as the only generator that changes in the presence of interactions because it generates dynamical evolution in time and interactions only change that property.

All the other generators imply purely geometric transformations which are not dynamical.

We also retain the definition $\hat{\mathbf{V}} = i[\hat{H}, \hat{\mathbf{Q}}]$, but note that interactions (a change in $\hat{H}$) will affect its value.

So we assume that $\hat{H}$ still satisfies

$$\hat{\mathbf{V}} = i[\hat{H}, \hat{\mathbf{Q}}]$$

If we shift to a different frame of reference moving uniformly with respect to the original frame, we saw earlier that $\hat{\mathbf{V}}$ transforms as

$$e^{i\vec{v}\cdot\hat{\mathbf{G}}}\hat{\mathbf{V}}e^{-i\vec{v}\cdot\hat{\mathbf{G}}} = \hat{\mathbf{V}} - \vec{v}\hat{I}$$

Expanding the left-hand side to 1st-order in $\vec{v}$ gives

$$i[\vec{v}\cdot\hat{\mathbf{G}},\hat{\mathbf{V}}] = -\vec{v}\hat{I} \text{ and } [\hat{G}_\alpha,\hat{P}_\beta] = i\delta_{\alpha\beta}M\hat{I}$$

Now, we still have the relation $\hat{G}_\alpha = M\hat{Q}_\alpha$, since its derivation does not use any commutators involving Ĥ.

Our earlier solution for $\hat{\mathbf{V}}$ (no external fields) was

$$\hat{\mathbf{V}} = \frac{\hat{\mathbf{P}}}{M}$$

Now the commutators

$$[\hat{G}_\alpha,\hat{V}_\beta] = i\delta_{\alpha\beta}\hat{I} \text{ and } [\hat{G}_\alpha,\hat{P}_\beta] = i\delta_{\alpha\beta}M\hat{I}$$

imply that $\hat{\mathbf{V}} = \hat{\mathbf{P}}/M$ commutes with $\hat{\mathbf{G}}$.

Since $\hat{\mathbf{G}} = M\hat{\mathbf{Q}}$, we must have

$$\hat{\mathbf{V}} - \frac{\hat{\mathbf{P}}}{M} = -\frac{\vec{A}(\hat{\mathbf{Q}})}{M} = \text{a function only of } \hat{\mathbf{Q}}$$

or

$$\hat{\mathbf{V}} = \frac{\hat{\mathbf{P}} - \vec{A}(\hat{\mathbf{Q}})}{M}$$

We now need to solve $\hat{\mathbf{V}}$ for $\hat{H}$.

We have the result

$$\frac{\hat{\mathbf{P}} - \vec{A}(\hat{\mathbf{Q}})}{M} = i[\hat{H},\hat{\mathbf{Q}}]$$

A possible solution is

$$\hat{H}_0 = \frac{\left(\hat{\mathbf{P}} - \vec{A}(\hat{\mathbf{Q}})\right)^2}{2M}$$

as can be seen from the derivation below.

We have

$$[\hat{H}_0, \hat{Q}_\alpha] = \frac{1}{2M}\sum_\beta \left[(\hat{P}_\beta^2 - A_\beta \hat{P}_\beta - \hat{P}_\beta A_\beta - A_\beta^2), \hat{Q}_\alpha\right]$$

$$= \frac{1}{2M}\sum_\beta \left([\hat{P}_\beta^2, \hat{Q}_\alpha] - [A_\beta \hat{P}_\beta, \hat{Q}_\alpha] - [\hat{P}_\beta A_\beta, \hat{Q}_\alpha] - [A_\beta^2, \hat{Q}_\alpha]\right)$$

Now

$$[\hat{Q}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I} \text{ and } [A_\beta, \hat{Q}_\alpha] = 0$$

which gives

$$[A_\beta \hat{P}_\beta, \hat{Q}_\alpha] = A_\beta[\hat{P}_\beta, \hat{Q}_\alpha] = -iA_\beta\delta_{\alpha\beta}\hat{I}$$

$$[\hat{P}_\beta A_\beta, \hat{Q}_\alpha] = [\hat{P}_\beta, \hat{Q}_\alpha]A_\beta = -iA_\beta\delta_{\alpha\beta}\hat{I}$$

$$[\hat{P}_\beta^2, \hat{Q}_\alpha] = \hat{P}_\beta\hat{P}_\beta\hat{Q}_\alpha - \hat{Q}_\alpha\hat{P}_\beta\hat{P}_\beta = \hat{P}_\beta[\hat{P}_\beta, \hat{Q}_\alpha] - i\delta_{\alpha\beta}\hat{P}_\beta = -2i\delta_{\alpha\beta}\hat{P}_\beta$$

$$[A_\beta^2, \hat{Q}_\alpha] = 0$$

We then have the final step in the proof

$$[\hat{H}_0, \hat{Q}_\alpha] = \frac{1}{2M}\sum_\beta 2i\delta_{\alpha\beta}(\hat{P}_\beta - A_\beta) = \frac{\hat{P}_\alpha - A_\alpha}{M}$$

That completes the proof.

Finally, we can then say

$$[(\hat{H} - \hat{H}_0), \hat{Q}_\alpha] = 0$$

which implies that at most $\hat{H}$ can differ from $\hat{H}_0$ only by a function of $\hat{\mathbf{Q}}$

$$\hat{H} - \hat{H}_0 = W(\hat{\mathbf{Q}}) = \text{ function of } \hat{\mathbf{Q}}$$

or, in general

$$\hat{H} = \frac{\left(\hat{\mathbf{P}} - \vec{A}(\hat{\mathbf{Q}})\right)^2}{2M} + W(\hat{\mathbf{Q}})$$

This is the only form of $\hat{H}$ **consistent with invariance under the Galilei group of transformations**.

The two new functions of $\hat{\mathbf{Q}}$ are called

$$\vec{A}(\hat{\mathbf{Q}}) = \text{vector potential and } W(\hat{\mathbf{Q}}) = \text{scalar potential}$$

Both of the functions can be time-dependent.

As operators they are functions only of $\hat{\mathbf{Q}}$ and not of $\hat{\mathbf{P}}$.

This form certainly includes the classical electromagnetic interaction.

However, we cannot identify $\vec{A}(\hat{\mathbf{Q}})$ and $W(\hat{\mathbf{Q}})$ with the electromagnetic potential because nothing in the derivation implies that they need to satisfy Maxwell's equations.

An Aside

The conventional notation involving the generators and the constant $\hbar$ is to use

$$\hat{\mathbf{P}} \to \frac{\hat{\mathbf{P}}}{\hbar} \quad , \quad \hat{\mathbf{J}} \to \frac{\hat{\mathbf{J}}}{\hbar} \quad , \quad M \to \frac{M}{\hbar} \quad , \quad \hat{H} \to \frac{\hat{H}}{\hbar}$$

leading to changed transformation operators of the form

$$e^{-i\vec{a}\cdot\hat{\mathbf{P}}/\hbar} \quad , \quad e^{-i\theta\mathbf{n}\cdot\hat{\mathbf{J}}/\hbar} \quad , \quad e^{-it\hat{H}/\hbar}$$

or we could just let $\hbar = 1$ and continue to use all previous results unchanged.

(1) The book next covers Multiparticle Systems, which we will skip over because we will not need this material for these lectures.

(2) The book covers a wide range of topics having to do with quantum measurement, entanglements, and so on, which we will skip over because we will not need this material for these lectures.

(3) The book covers many aspects of density operator theory, which we will skip over because we will not need this material for these lectures.

Equations of Motion Revisited and Finished

We discussed time dependence and the time evolution operator earlier.

This is the most important topic in quantum mechanics and will eventually enable us to make predictions about the behavior of real physical systems.

Let us now review our earlier discussion from a more general point of view in light of some of the new ideas we have introduced.

We derived a differential equation of motion for the state vector of the form

$$\frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}(t)|\psi(t)\rangle$$

If, initially (at $t = t_0$) we are in the state represented by the ket vector $|\psi(t_0)\rangle = |\psi_0\rangle$, then we wrote the formal solution of the differential equation in terms of the time development operator $\hat{U}(t, t_0)$ as

$$|\psi(t)\rangle = \hat{U}(t, t_0)\,|\psi(t_0)\rangle$$

We also showed that $\hat{U}(t, t_0)$ satisfies the same differential equation as $|\psi(t)\rangle$

$$\frac{\partial}{\partial t}\hat{U}(t, t_0) = -\frac{i}{\hbar}\hat{H}(t)\hat{U}(t, t_0)$$

with the boundary condition $\hat{U}(t_0, t_0) = \hat{I}$.

This implies, using the relation $(\hat{A}\hat{B})^\dagger = \hat{B}^\dagger\hat{A}^\dagger$ that

$$\frac{\partial}{\partial t}\hat{U}^\dagger(t, t_0) = \frac{i}{\hbar}\hat{U}^\dagger(t, t_0)\hat{H}(t)$$

We then have

$$\begin{aligned}\frac{\partial}{\partial t}(\hat{U}^\dagger\hat{U}) &= \hat{U}^\dagger\frac{\partial\hat{U}}{\partial t} + \frac{\partial\hat{U}^\dagger}{\partial t}\hat{U} \\ &= -\frac{i}{\hbar}\hat{U}^\dagger\hat{H}\hat{U} + \frac{i}{\hbar}\hat{U}^\dagger\hat{H}^\dagger\hat{U} \\ &= \frac{i}{\hbar}\hat{U}^\dagger(\hat{H} - \hat{H}^\dagger)\hat{U}\end{aligned}$$

If $\hat{H}$ is Hermitian, then $\hat{H} = \hat{H}^\dagger$ and

$$\frac{\partial}{\partial t}(\hat{U}^\dagger\hat{U}) = 0$$

which implies that

$$\hat{U}^\dagger(t, t_0)\hat{U}(t, t_0) = c\hat{I} \text{ where } c = \text{constant}$$

But the boundary condition implies that which implies that $c = 1$.

Therefore, in general we have $\hat{U}^\dagger(t,t_0)\hat{U}(t,t_0) = \hat{I}$, which is the **unitarity condition** and implies $\hat{U}^\dagger = \hat{U}^{-1}$ or that $\hat{U}$ is unitary when $\hat{H}$ is Hermitian.

This agrees with our earlier results.

Now, if $\hat{H}(t)$ is independent of t, then the differential equation has a simple solution

$$\hat{U}(t,t_0) = e^{-i(t-t_0)\frac{\hat{H}}{\hbar}}$$

which is the form we assumed for the time evolution operator during the discussion of symmetry transformations.

If $\hat{H}(t)$ is not independent of t, then no simple closed form solution can be given for $\hat{U}(t, t_0)$.

We must use perturbation theory as we shall see later when we develop time-dependent perturbation theory.

We can now derive the **equation of motion for the density operator**.

We use a pure state for simplicity.

We have

$$\begin{aligned}\hat{W}(t) &= |\psi(t)\rangle\langle\psi(t)| = \hat{U}(t,t_0)\,|\psi(t_0)\rangle\langle\psi(t_0)|\,\hat{U}^\dagger(t,t_0)\\ &= \hat{U}(t,t_0)\hat{W}(t_0)\hat{U}^\dagger(t,t_0)\end{aligned}$$

Now, using $\hat{W}(t_0) = \hat{W}_0$, differentiating with respect to t gives

$$\frac{d\hat{W}(t)}{dt} = \frac{d\hat{U}(t,t_0)}{dt}\hat{W}_0\hat{U}^\dagger(t,t_0) + \hat{U}(t,t_0)\hat{W}_0\frac{d\hat{U}^\dagger(t,t_0)}{dt}$$
$$= -\frac{i}{\hbar}\hat{H}\hat{U}(t,t_0)\hat{W}_0\hat{U}^\dagger(t,t_0) + \frac{i}{\hbar}\hat{U}(t,t_0)\hat{W}_0\hat{U}^\dagger(t,t_0)\hat{H}$$
$$= -\frac{i}{\hbar}[\hat{H}(t),\hat{W}(t)]$$

We will assume that this equation is also true for general states.

Now, we have stated earlier that no physical significance can be attached in quantum mechanics to operators and vectors.

The only physically significant objects are the probability distributions of observables or their expectation values.

Earlier we derived the result

$$\langle\hat{Q}\rangle = Tr(\hat{W}\hat{Q})$$

We now assume that this result carries over to the time-dependent case and we have

$$\langle\hat{Q}\rangle_t = Tr(\hat{W}(t)\hat{Q})$$

Using the above expression for $\hat{W}(t)$ and the fact that the trace is invariant under cyclic permutation, i.e.,

$$Tr(\hat{A}\hat{B}\hat{C}) = Tr(\hat{C}\hat{A}\hat{B}) = Tr(\hat{B}\hat{C}\hat{A})$$

we get

$$\langle\hat{Q}\rangle_t = Tr(\hat{W}\hat{Q}) = Tr(\hat{U}(t,t_0)\hat{W}_0\hat{U}^\dagger(t,t_0)\hat{Q})$$
$$= Tr(\hat{W}_0\hat{U}^\dagger(t,t_0)\hat{Q}\hat{U}(t,t_0))$$

This result is the formal basis of the Schrodinger and Heisenberg pictures we discussed earlier.

If we leave the time dependence in the density operator, then

$$\langle \hat{Q} \rangle_t = Tr(\hat{W}(t)\hat{Q})$$

where

$$\frac{d\hat{W}(t)}{dt} = -\frac{i}{\hbar}[\hat{H}(t), \hat{W}(t)] \text{ and } \frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}(t)|\psi(t)\rangle$$

$\hat{Q}$ is independent of time is these equations.

This is the Schrodinger picture.

On the other hand, if we write

$$\langle \hat{Q} \rangle_t = Tr(\hat{W}_0 \hat{U}^{\dagger}(t, t_0)\hat{Q}\hat{U}(t, t_0)) = Tr(\hat{W}_0 \hat{Q}_H(t))$$

where

$$\hat{Q}_H(t) = \hat{U}^{\dagger}(t, t_0)\hat{Q}\hat{U}(t, t_0)$$

then the operator is time dependent and the density operator(and hence the state vectors) are independent of time.

This is the Heisenberg picture.

We can derive the equation of motion of the time-dependent operators in the Heisenberg picture as follows:

$$\frac{d}{dt}\hat{Q}_H(t) = \frac{\partial \hat{U}^\dagger}{\partial t}\hat{Q}\hat{U} + \hat{U}^\dagger\frac{\partial \hat{Q}}{\partial t}\hat{U} + \hat{U}^\dagger\hat{Q}\frac{\partial \hat{U}}{\partial t}$$

$$= \frac{i}{\hbar}(\hat{U}^\dagger\hat{H}\hat{Q}\hat{U} - \hat{U}^\dagger\hat{Q}\hat{H}\hat{U}) + \hat{U}^\dagger\frac{\partial \hat{Q}}{\partial t}\hat{U}$$

$$= \frac{i}{\hbar}(\hat{U}^\dagger\hat{H}\hat{U}\hat{U}^\dagger\hat{Q}\hat{U} - \hat{U}^\dagger\hat{Q}\hat{U}\hat{U}^\dagger\hat{H}\hat{U}) + \hat{U}^\dagger\frac{\partial \hat{Q}}{\partial t}\hat{U}$$

$$= \frac{i}{\hbar}[\hat{H}_H(t), \hat{Q}_H(t)] + \left(\frac{\partial \hat{Q}}{\partial t}\right)_H$$

where we have included the possibility that $\hat{Q}$ has some explicit time dependence and we have used the definition

$$\hat{A}_H(t) = \hat{U}^\dagger(t, t_0)\hat{A}\hat{U}(t, t_0)$$

for any operator.

Note the change in sign in front of the commutator in this equation for the operator from that of the density operator in the Schrodinger picture.

The two pictures are clearly equivalent as mathematical formalisms, since they are derived from the same expectation value formula by an internal rearrangement of terms.

Another way to say this is that the two pictures are equivalent because the only physically significant quantity $\langle\ \hat{Q}\ \rangle_t$ depends only on the relative motion in time of $\hat{W}$ and $\hat{Q}$, which is the same in both cases.

In the Schrodinger picture $\hat{W}(t)$ moves forward in time (term of the form $\hat{U}\hat{W}_0\hat{U}^\dagger = \hat{U}\hat{W}_0\hat{U}^{-1}$) and in the Heisenberg picture $\hat{Q}(t)$ moves backward in time (term of the form $\hat{U}^\dagger\hat{W}_0\hat{U} = \hat{U}^{-1}\hat{W}_0\hat{U}$)

The opposite senses of motion in time produce the sign difference of the commutator terms in the equations of motion.

These two picture are mutually exclusive and cannot be used together.

Finally, we determine expressions for $d\langle\hat{Q}\rangle_t/dt$ in each picture.

In the Schrodinger picture

$$\frac{d\langle\hat{Q}\rangle_t}{dt} = \frac{d}{dt}Tr(\hat{W}(t)\hat{Q}) = Tr\left[\frac{d\hat{W}}{dt}\hat{Q} + \hat{W}\frac{\partial\hat{Q}}{\partial t}\right]$$

$$= Tr\left[-\frac{i}{\hbar}(\hat{H}\hat{W}\hat{Q} - \hat{W}\hat{H}\hat{Q}) + \hat{W}\frac{\partial\hat{Q}}{\partial t}\right]$$

$$= Tr\left[-\frac{i}{\hbar}(\hat{W}\hat{Q}\hat{H} - \hat{W}\hat{H}\hat{Q}) + \hat{W}\frac{\partial\hat{Q}}{\partial t}\right]$$

$$= Tr\left[\frac{i}{\hbar}\hat{W}(t)[\hat{H},\hat{Q}] + \hat{W}\frac{\partial\hat{Q}}{\partial t}\right]$$

In the Heisenberg picture

$$\frac{d\langle\hat{Q}\rangle_t}{dt} = Tr(\hat{W}_0\frac{d\hat{Q}_H(t)}{dt}) = Tr\left[\frac{i}{\hbar}\hat{W}_0[\hat{H},\hat{Q}_H(t)] + \hat{W}_0\left(\frac{\partial\hat{Q}}{\partial t}\right)_H\right]$$

For a pure state, we can rewrite these results in terms of the state vectors instead of the density operator.

We have in the Schrodinger picture

$$\langle \hat{Q} \rangle_t = \langle \psi(t) | \hat{Q} | \psi(t) \rangle \text{ where } |\psi(t)\rangle = \hat{U}(t, t_0) |\psi_0\rangle$$

and in the Heisenberg picture

$$\langle \hat{Q} \rangle_t = \langle \psi_0 | \hat{Q}_H(t) | \psi_0 \rangle \text{ where } \hat{Q}_H(t) = \hat{U}^\dagger(t, t_0) \hat{Q} \hat{U}(t, t_0)$$

as we saw in our earlier discussions.

Symmetries, Conservation Laws and Stationary States

Let $\hat{T}(t) = e^{is\hat{K}}$ represent a continuous unitary transformation with a Hermitian generator $\hat{K} = \hat{K}^\dagger$.

Another operator $\hat{A}$ representing some observable is invariant under this transformation if

$$\hat{T}(s) \hat{A} \hat{T}^{-1}(s) = \hat{A}$$

or

$$\hat{A}\hat{T}(s) - \hat{T}(s)\hat{A} = [\hat{A}, \hat{T}(s)] = 0$$

If s is an infinitesimal, then we can write

$$(\hat{I} + is\hat{K}) \hat{A} (\hat{I} - is\hat{K}) = \hat{A} \rightarrow \hat{A} + is[\hat{K}, \hat{A}] = \hat{A}$$

or

$$[\hat{K}, \hat{A}] = 0$$

In words, the invariance of $\hat{A}$ under the continuous transformation $\hat{T}(t) = e^{isK}$ for all s implies it is true for infinitesimal *s* and thus leads to the commutator condition for invariance, which says that the operator commutes with the Hermitian generator of the transformation.

It works both ways:

$$[\hat{T}, \hat{A}] = 0 \rightarrow \text{ invariance under finite transformation}$$
$$\rightarrow \text{ invariance under infinitesimal transformation}$$
$$\rightarrow [\hat{K}, \hat{A}] = 0$$

or

$$[\hat{K}, \hat{A}] = 0 \rightarrow \text{ invariance under infinitesimal transformation}$$
$$\rightarrow \text{ invariance under finite transformation}$$
$$\rightarrow [\hat{T}, \hat{A}] = 0$$

If $\hat{K}$ depends on *t*, then the commutators $[\ \hat{K}(t), \hat{A}\ ] = 0$ and $[\ \hat{T}(t), \hat{A}\ ] = 0$ must hold for all *t*.

Now, the Hermitian generators of the symmetry transformation, as we have seen, correspond to dynamical variables of a physical system.

$$\text{space displacements} \Leftrightarrow \hat{\mathbf{P}}$$
$$\text{rotations} \Leftrightarrow \hat{\mathbf{J}}$$
$$t \text{ displacements} \Leftrightarrow \hat{H}$$

These symmetry generators have no explicit time dependence, i.e., $\partial \hat{K}/\partial t = 0$.

Therefore,

$$\frac{d\langle \hat{K} \rangle_t}{dt} = Tr\left[\frac{i}{\hbar}\hat{W}(t)[\hat{H},\hat{K}]\right] = 0$$

If $\hat{H}$ is invariant under the corresponding symmetry transformation, i.e., $[\hat{H}, \hat{K}] = 0$.

Now $[\hat{H}, \hat{K}] = 0 \rightarrow [\hat{H}, f(\hat{K})] = 0$.

This says that we must have

$$[\hat{H}, \theta(x - \hat{K})] = 0$$

But in our probability discussions, we showed that

$$\langle \theta(x - \hat{K}) \rangle = Prob(K < x|\hat{W})$$

= probability that observable $\hat{K}$ has a value $< x$ given $\hat{W}$

This then says that $Prob(K < x|\hat{W})$ is independent of t no matter what initial state we start with.

In this case, observable K = a is a constant of the motion.

Examples

$[\hat{H}, \hat{P}_\alpha] = 0$ implies invariance under a space displacement along the α-axis.

This implies that $\hat{P}_\alpha$ = constant of the motion (is conserved) = linear momentum along the α-axis.

$[\hat{H}, \hat{J}_\alpha] = 0$ implies invariance under a rotation about the α-axis.

This implies α that $\hat{J}_\alpha$= constant of the motion (is conserved) = angular momentum about α the α-axis.

If $\hat{H}$ is not an explicit function of t, then since $[\hat{H}, \hat{H}] = 0$, $\hat{H}$ is invariant under time translations = constant of the motion = energy of the system.

Now suppose that $\hat{H}$ is independent of t and that

$$|\psi(0)\rangle = [\text{ eigenvector of } \hat{H}] = |E_n\rangle \text{ such that } \hat{H}|E_n\rangle = E_n|E_n\rangle$$

Then we have, using $[\hat{H}, \hat{U}] = [\hat{H}, f(\hat{U})] = 0$,

$$\begin{aligned}
\frac{d}{dt}|\psi(t)\rangle &= -\frac{i}{\hbar}\hat{H}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}\hat{U}(t,0)|\psi(0)\rangle \\
&= -\frac{i}{\hbar}\hat{H}\hat{U}(t,0)|E_n\rangle = -\frac{i}{\hbar}\hat{H}\hat{U}(t,0)\hat{H}|E_n\rangle \\
&= -\frac{i}{\hbar}\hat{H}\hat{U}(t,0)E_n|E_n\rangle = -\frac{i}{\hbar}E_n\hat{H}\hat{U}(t,0)|E_n\rangle \\
&= -\frac{i}{\hbar}E_n|\psi(t)\rangle
\end{aligned}$$

which has the solution

$$|\psi(t)\rangle = e^{-i\frac{E_n}{\hbar}t}|E_n\rangle$$

In this case, we then have for the expectation value of any observable represented by the operator $\hat{R}$

$$\langle\hat{R}\rangle = \langle\psi(t)|\hat{R}|\psi(t)\rangle = \langle E_n|e^{i\frac{E_n}{\hbar}t}\hat{R}e^{-i\frac{E_n}{\hbar}t}|E_n\rangle = \langle E_n|\hat{R}|E_n\rangle$$

or $\langle \hat{R} \rangle$ is independent of t (for this state).

This implies that $\langle f(\hat{R})\rangle$ is also independent of t, which finally, implies that $\langle\theta(x - \hat{R})\rangle$ is independent of t.

This means that, in this state,

$$Prob(x < R|\psi) \text{ is independent of } t$$

This kind of state is called a **stationary state**.

In a stationary state the expectation values and probabilities of all observables are independent of time.

If an observable is a constant of the motion, however, then this would be true for any state and not just a stationary state.

So these are very different physical concepts.

Now, if $[\hat{K}, \hat{H}] = 0$, then $\hat{K}$ and $\hat{H}$ have a common set of eigenvectors (it is a complete set).

But the eigenvectors of $\hat{H}$ are stationary states.

This means that we can prepare systems in stationary states where both the energy and the observable represented by $\hat{K}$ have definite values (no dispersion).

Suppose we have a set of mutually commuting observables and that they all also commute with $\hat{H}$.

Then they have a common set of eigenvectors.

We will use the eigenvalues of $\hat{H}$ and all the eigenvalues of this mutually commuting set of observables to label state vectors.

We will call the labels quantum numbers.

They will designate all we know about a state vector.

Chapter 5 - How Does It really Work: Photons, K-Mesons and Stern-Gerlach

Introduction

Many experiments indicate that electromagnetic waves have a vector property called *polarization*.

Suppose that we have an electromagnetic wave (we will just say light from now on) passing through a piece of Polaroid material.

The Polaroid material has the property that it lets through only that light whose polarization vector is oriented parallel to a preferred direction in the Polaroid(called the optic axis).

Classically, if an incident beam of light is polarized parallel to the optic axis, then experiment says that all of its energy gets through the Polaroid.

If, on the other hand, it is polarized perpendicular to the optic axis, then experiment says that none of its energy gets through the Polaroid.

Classically, if it is polarized at an angle α to the optic axis, then experiment says that a *fraction* $\cos^2\alpha$ of its energy gets through the Polaroid.

Many experiments, including several that we will discuss later, indicate that light(and everything else in the universe as it turns out) exhibits both "particle- like" and "wave-like" properties.

We will define both of these terms more carefully later and decide then if this simple statement makes any sense.

The particle associated with light is called a photon.

Can the experimental polarization/Polaroid results be reconciled with this idea of a particle-like photon?

Suppose we assume that a light wave that is polarized in a certain direction is made up of a large number of photons each of which is polarized in that same direction.

The particle properties of light, as represented by photons, invariably lead to some confusion.

It is not possible to eliminate all of this confusion at this elementary discussion level because a satisfactory treatment of photons requires quantum electrodynamics.

We can, however, make many of the more important physical properties clear.

Consider a simple representation of a monochromatic electromagnetic wave (and its associated photons) with angular frequency ω and wavelength λ moving in a direction given by the unit vector $\hat{k}$.

Such a monochromatic electromagnetic wave is composed of N (a very large number) photons, each with energy E and momentum $\vec{p}$, such that we have the relationships

$$E = \hbar\omega \quad , \quad \vec{p} = \hbar\vec{k} = \frac{h}{\lambda}\hat{k}$$

where $\vec{k}$ is the wave vector, $\hbar = h/2\pi$, $\omega = 2\pi f$, h = Planck's constant, f = frequency = c/λ, and c = speed of light.

We note that

$$E = \frac{h}{2\pi}2\pi f = hf = h\frac{c}{\lambda} = pc$$

as required by relativity for a particle with zero mass(such as the photon).

The number of photons in the wave is such that the total energy of the N photons, $NE = N\hbar\omega$, is equal to the total energy W in the electromagnetic wave, i.e., $W = NE = N\hbar\omega$.

Here, we are using the fact, derived from many experiments, that the energy of the light wave is quantized and thus can only take on certain discrete values (its value is a multiple of some quantum of energy).

When we specify the polarization of light, we are actually giving the direction of the electric field vector $\vec{E}$.

Light waves are generally represented by plane electromagnetic waves.

This means that the electric field vector $\vec{E}$ and the magnetic field vector $\vec{B}$ are both perpendicular to the direction of propagation specified by $\vec{k}$.

If we choose the direction of propagation to be the z-axis, which is specified by the unit vector $\hat{e}_z$, then $\vec{E}$ and $\vec{B}$ lie in the x–y plane.

Since we are only considering the polarization property at this point, we can concentrate on the $\vec{E}$ vector alone.

Now, any vector in the x–y plane can be specified in terms of a pairs of orthonormal vectors (called the basis) in that plane.

The vectors in these orthonormal directions are called the **polarization** vectors.

Two standard sets of orthonormal polarization vectors are often chosen when one discusses polarization.

One of the two sets is

$$\hat{e}_x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad \hat{e}_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

which correspond to plane(or **linearly**)-polarized waves.

A second set is

$$\hat{e}_R = \hat{e}_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} \quad , \quad \hat{e}_L = \hat{e}_- = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix}$$

which correspond to circularly polarized waves.

For classical electromagnetic fields, a light wave propagating in the z-direction is usually described(illustrating with the two orthonormal sets) by electric field vectors of the forms given below.

Plane-polarized basis:

$$\vec{E}(\vec{r},t) = \begin{pmatrix} E_x(\vec{r},t) \\ E_x(\vec{r},t) \\ 0 \end{pmatrix} = E_x(\vec{r},t)\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + E_y(\vec{r},t)\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$= E_x(\vec{r},t)\hat{e}_x + E_y(\vec{r},t)\hat{e}_y$$

Circularly-polarized basis:

$$\vec{E}(\vec{r},t) = \begin{pmatrix} E_x(\vec{r},t) \\ E_x(\vec{r},t) \\ 0 \end{pmatrix}$$

$$= \frac{E_x(\vec{r},t) + iE_y(\vec{r},t)}{\sqrt{2}}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} + \frac{E_x(\vec{r},t) - iE_y(\vec{r},t)}{\sqrt{2}}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix}$$

$$= E_x(\vec{r},t)\hat{e}_R + E_y(\vec{r},t)\hat{e}_L$$

By convention, we represent the field components by

$$E_x(\vec{r},t) = E_x^0 e^{i(kz-\omega t+\alpha_x)} \text{ and } E_y(\vec{r},t) = E_y^0 e^{i(kz-\omega t+\alpha_y)}$$

where α_x and α_y are the (real)phases and E_x^0 and E_y^0 are the (real) amplitudes of the electric field components.

Clearly, the polarization state of the light is directly related to the $\vec{E}$ vectors in this formulation.

For example, using the above equations, we have these cases:

1. If $E_y = 0$, then the wave is *plane polarized* in the x-direction

$$\vec{E} = E_x \hat{e}_x = E_x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

2. If $E_x = 0$, then the wave is *plane polarized* in the y-direction

$$\vec{E} = E_y \hat{e}_y = E_y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

3. If $E_x = E_y$, then the wave is *plane polarized* at 45°

$$\vec{E} = E_x \hat{e}_x + E_y \hat{e}_y = E_x \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

4. If $E_y = -iE_x = e^{-i\pi/2} E_x$, then the y-component lags the x-component by 90° and the wave is *right circularly* polarized

$$\vec{E} = E_x \hat{e}_R = E_x \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix}$$

5. If $E_y = iE_x = e^{i\pi/2} E_x$, then the y-component leads the x-component by 90° and the wave is *left circularly* polarized

$$\vec{E} = E_x \hat{e}_L = E_x \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix}$$

For our present discussion, the above properties of polarization are sufficient.

We will give a more detailed discussion of the quantum mechanics of photon polarization shortly.

The picture we are proposing assumes that each photon has the same polarization as the light wave, which is, in fact, verified by experiment.

This simple experimental property leads to some fundamental difficulties for classical mechanics.

If the incident beam is polarized parallel or perpendicular to the optic axis of a Polaroid, then classical physics has no problems ..... all the photons (and thus all the energy) either pass through or do not pass(and thus none of the energy) through the Polaroid.

But what about the case where the wave is polarized at 45° to the optic axis of the Polaroid?

For the beam as a whole, the experimental result is that 1/2 ($\cos^2 45^\circ = 1/2$) of the total energy and hence 1/2 of the photons pass through.

But what about **any particular photon**, each of which is polarized at 45° to the optic axis?

Now the answer is not clear at all and the fundamental dilemma of the sub-atomic world rears its ugly head.

As will become clear during our discussions of quantum mechanics, this question about what will happen to a particular photon under certain conditions is not very precise.

In order for any theory to make clear predictions about experiments, we will have to learn how to ask very precise questions.

We must also remember that only questions about the results of experiments have a real significance in physics and it is only such questions that theoretical physics must consider.

All relevant questions and the subsequent experiments devised to answer the questions must be clear and precise, however.

In this case, we can make the question clear by doing the experiment with a beam containing only one photon and observe what happens after it arrives at the Polaroid.

In particular, we make a simple observation to see whether or not it passes through the Polaroid.

The most important experimental result is that this single photon either passes through the Polaroid or it does not.

I will call this type of experiment a *go-nogo* or *yes-no* experiment.

We never observe 1/2 the energy of a single photon.

We always observe either zero energy or an energy exactly equal to $\hbar\,\omega$.

One never observes a part of a photon passing through and a part getting absorbed in the Polaroid.

In addition, if a photon gets through, then experiment says that its polarization vector changes such that it ends up polarized in a direction parallel to the optic axis of this particular Polaroid (instead of at 45° with respect to that axis as it was polarized beforehand).

In a beam of *N* photons, each photon will independently behave as the single photon did in the description above.

No experiment can determine which of the photons will pass through and which will not, even though they are all identical.

In each experiment, however, exactly 1/2 of the total energy and 1/2 of the photons will pass through the 45° Polaroid.

As we shall show later, the only way this result can be interpreted is to say that each photon has a probability = 1/2 of passing through the 45° Polaroid.

We are led to this probabilistic point of view because the energy of the electromagnetic wave is quantized(or equivalently, that the electromagnetic wave is made up of photons) and we cannot have fractions of the energy quantum appearing during an experiment.

We have managed to preserve the indivisibility of the photons(the quantization of their energy), but we were able to do this only by abandoning the comforting determinacy of classical physics and introducing probability.

The results in this experiment are not completely determined by the experimental conditions(initial) under control of the experimenter, as they would have been according to classical ideas.

As we shall see, the most that we will be able to predict in any experiment is a set of possible results, with a probability of occurrence for each.

The experiment described above involving a single photon polarized at an angle to the optic axis, allows us to ask only one type of experimental and theoretical question, namely, does the photon go through or is it absorbed - only a *yes-no* experiment?

That will turn out to be the only legitimate question we can ask in this case.

It is the first indication of the way we should frame our discussion of theory!

We shall see that questions like....

What decides whether a particular photon goes through?

When does a particular photon decide whether it will pass through?

How does a particular photon change its polarization direction?

cannot be answered by experiment and, therefore, must be regarded as *outside the domain* of quantum theory and possibly all of physics.

What will our theory of quantum mechanics say about the state of the single photon?

It will be shown that the photon polarized at an angle to the optic axis is in a very special kind of state that we will call a *superposition* of being polarized perpendicular to the optic axis and of being polarized parallel to the optic axis, i.e., *a superposition of all the possibilities*.

In this state, there will exist an extraordinary kind of relationship between the two kinds(mutually perpendicular directions) of polarization.

The meaning of the word *superposition* will follow clearly from the mathematical formalism and language we have developed in these lectures.

It will, however, require a new physical connection to mathematics.

This, as we shall see later, is suggested by an attempt to express the meaning of superposition in ordinary language(words).

If we attempt to explain the behavior of the photon polarized at an angle to the optic axis using ordinary language, then we might be inclined to say something like this:

The photon is

not polarized *parallel* to the optic axis

not polarized *perpendicular* to the optic axis

not simultaneously possessing *both* polarizations

not possessing *neither* polarization

For this experiment with only two possible polarizations, these statements exhaust all the logical possibilities allowed by *ordinary words*!

Superposition is something *completely* different than any of the above and it is *not all* of the above.

Its physical content will, however, be precise and clear in our new mathematical formalism.

When the photon encounters the Polaroid, we are observing it.

We are observing whether it is polarized perpendicular or parallel to the optic axis of the Polaroid.

The effect of this measurement will be to end up with the photon having one or the other polarizations(the one we measure).

In such a measurement, the photon always makes a jump from a state of superposition to a state of a definite polarization.

Which of the two states it jumps to cannot be predicted.

We will, as we shall see, be able to predict the probability of each possibility.

If it jumps into the parallel state, it has passed through.

If it jumps into the perpendicular state, it has been absorbed.

We will have a great deal more to say about the two new words, *superposition* and *jump*, as we proceed.

Do not attach any classical meaning to the word jump as used here.

It simply my attempt to use words for the moment.

Photon Interference

Another classic experiment involving light *waves* and hence photons is two-slit interference.

The ideas of *superposition* and *jump* will reappear in this case and give us another view to enhance our understanding(or maybe confusion) at this point.

We will discuss this type of experiment in great detail and in many different forms in these lectures.

As we shall see, the meaning of this experiment is easily misunderstood.

This experiment looks at the position and momentum properties of photons instead of the polarization properties.

For an approximately monochromatic light wave we have some knowledge of both the position and momentum of the photons.

In particular, the photons must be located near the beam of light and their momentum is in the direction of propagation of the beam and has an approximate magnitude

$$p = \frac{E}{c} = \frac{\hbar\omega}{c} = \frac{hf}{c} = \frac{h}{\lambda}$$

According to the standard wave interference description, a two-slit interference experiment consists of a device which, in some manner, splits the incoming beam into two beams which are sent over different physical paths and then re-combined, in some way, so that they can interfere and produce a distinctive pattern on a screen.

To see the problems directly, we again consider an incident beam consisting of a **single photon**.

What happens as it passes through the apparatus?

The photon has only two classical possibilities if it is going to reach the screen and contribute to the interference pattern, i.e., it must follow a path that passes through one of the two slits.

As in the polarization experiment, the correct quantum description will involve describing the single photon as a *superposition* of photons traveling on (at least) two different paths to the same point on a screen.

Once again, as we shall see in detail later on, if we tried to use ordinary language to describe the experimental results in the interference experiment we find ourselves saying that the photon is

not on path #1

not on path #2

not simultaneously on *both* paths

not on *neither* path

What actually happens when we try to determine the energy of a photon in one of the beams?

When we do the measurement, we always find the photon(all of the energy in the single photon system) in one of the beams.

When we observe the photon it must somehow *jump* from being in a superposition (which generates the interference pattern) to being entirely in one of the beams (which, as it turns out, does not generate the interference pattern).

One possibility is that the jump is caused by the measurement of the energy in this case.

Although, we cannot predict on which path the photon will be found, we can, however, predict the probability of either result from the mathematical formalism we will develop for superposition.

Even more striking is the following result.

Suppose we have a single beam of light consisting of a large number of photons, which we split up into two beams of equal intensity.

On the average we should have about half of the photons in each beam.

Now make the two beams interfere.

If we were to assume that the interference is the result of a photon in one beam interfering with a photon in the other beam, then we would have to allow two photons to annihilate each other some of the time (producing interference minima) and sometimes turn into four photons (producing interference maxima).

This process clearly violates conservation of energy.

The new concept of superposition will enable us to deal with this problem.

Each photon in the original beam will be in a superposition of being in both beams.

The superposition of possibilities will produce the interference effect.

Interference between two different photons or of the photon with itself is never required and energy always be conserved.

We will have a great deal more to say about these and other experiments as we develop the mathematical formalism of quantum mechanics.

Some questions about this approach

Is it really necessary to introduce the new concepts of superposition and jump?

In many simple experiments, the classical model of waves and photons connected in some vague statistical way will be sufficient to explain many experimental results.

The new quantum ideas do not contribute anything new in these cases.

As we shall see as we progress through this chapter, there are, however, an overwhelming number of experiments where the only correct explanation comes from a quantum theory built around these new concepts and their associated mathematical formalism.

Will this new theory give us a better model of the photon and of single photon processes?

I do not believe that the object of a physical theory is to provide a satisfying model or picture, but instead, to develop a language that enables us to formulate physical laws and make experimental predictions.

Models and pictures are holdovers from classical physics as applied to macroscopic phenomena.

In the case of atomic phenomena and beyond we cannot expect to find an appropriate model or picture.

All such models and pictures rely on classical concepts and these concepts are totally inadequate in the microscopic arena.

Their main goal seems to be to make the reader feel more comfortable with the new and strange phenomena under discussion.

Their many misleading implications, however, often cause misunderstandings that lead many a student down dead-end paths as they try to understand quantum mechanics.

We will avoid all non-mathematical models.

Unless one has mastered the correct language and adjusted one's mode of thinking, it will not, in general, be possible to understand modern theoretical physics.

For me, a self-consistent mathematical formulation of the theory that uses a language appropriate to the physical phenomena involved and is able to make correct predictions for all experiments constitutes a valid physical model.

The reasons for observed physical behavior are contained in the mathematical formalism.

There is no deeper physics involved beyond the mathematical formalism.

The only model is the mathematical formalism and its associated physical connections along with a language that is needed to understand it and express what it means in terms of experiments.

What about determinism?

The problems that the concept of superposition presents to the classical idea of determinism when we create a well-defined system are so devastating that these old ideas must be abandoned.

We will discuss these ideas in great detail throughout these lectures, but for now a short digression about states, superposition and measurements seems appropriate.

A classical state is specified by giving numerical values for all the coordinates and velocities of the constituents of the system at some instant of time.

The subsequent motion is then determined if we know the force laws.

For a small system, however, we cannot observe it and determine its properties to the extent required for a complete classical description.

Thus, any microscopic system is necessarily specified by fewer or more indefinite data than all the coordinates and velocities at some instant of time.

In the new mathematical formalism, we will carefully define the kind of information we can know about a state, which information we can actually know at a given instant of time and how we will prepare such states.

The prepared states will have definite values of a specific set of measurable quantities.

In terms of such states, we will define a superposition such that we have well-defined mathematical relationships between the different states making up the superposition.

The mathematical definition of a superposition will not be describable with the classical ideas or pictures or models available, i.e., it will turn out that when we superpose two states, the nature of the relationship that exists between them cannot be explained in terms of classical physical concepts.

The system is not partly in one state and partly in the other in any sense.

We will have to deal with a completely new idea here.

During the course of our discussions, we will need to get accustomed to the mathematical formalism behind the concept of superposition and we will need to rely on the formalism and the mathematics, as expressed by the Dirac lan- guage, without having any detailed classical models.

The new superposed state will be completely defined by the states involved, the mathematical relationships between them and the physical meaning of those mathematical relationships as defined by the formalism.

As we saw in the polarization and interference experiments, we must have a superposition of two possible states, say A and B.

Suppose that if the system were in state A alone, then a particular measurement would, with certainty, produce the result a and if the system were in state B alone, then the same measurement would, with certainty, produce the result b.

Then, if the system is in the superposed state(of A and B), it turns out that the same measurement will sometimes produce the result a and sometimes the result b.

Given the mathematical definition of the superposed state and the rules that we will specify to relate the physics to the mathematics, we will be able to predict the probability of getting the result a and the probability of getting the result b.

We will not be able to predict which result we will get in any one measurement, i.e., the measurement process will not be deterministic.

Identical measurements on identically prepared states will not yield identical results in any one measurement.

If we repeat the measurements a large number of times, however, we can predict the fraction of time we will obtain the result .

That is all we will be able to predict using the formalism of quantum mechanics we will develop.

I firmly believe, that nature is such that we will not, under any circumstances, be able to make any more detailed predictions for quantum systems.

So, summarizing the situation before we proceed.....

When physicists try to construct a theory from their experimental observations they are primarily concerned with two things:

how to calculate something that will enable them to predict the results of further experiments

how to understand what is going on in the experiments

Now, it is not always possible to satisfy both of these concerns at the same time.

Sometimes we have a reasonably clear idea of what is going on, but the mathematical details are so complex that calculating is very hard.

Sometimes the mathematics is quite clear, but the understanding is difficult.

Quantum mechanics falls into the latter category.

Over the years since quantum mechanics was first proposed, a very pragmatic attitude has formed.

Physicists realized that an understanding of how electrons could behave in such a manner would come once they developed the rules that would enable them to calculate the way they behaved.

So quantum mechanics as a set of rules was developed, as we shall see in these lectures.

It has been tested in a large number of experiments and never been found to be wrong.

It works extremely well.

Yet, we are still not sure how the electrons can behave in such a manner.

In many ways, the situation has even gotten worse over the years.

Some people, as we shall see, held out the hope that quantum mechanics was an incomplete theory and that as we did more experiments we would discover some loose end or new idea that would allow us to make sense of things.

This has not happened and I believe it never will.

We now carry out the details of a special case that will illustrate how Quantum Mechanics works and also illustrate the mathematical formalism that we have developed in earlier lectures.

Photon Polarization

As we mentioned earlier, the electric field vector $\vec{E}$ of plane electromagnetic waves lies in a plane perpendicular to the direction of propagation of the wave.

If we choose the z-axis as the direction of propagation, we can represent the electric field vector as a 2-dimensional vector in the $x - y$ plane.

This means that we will only require two numbers to describe the electric field.

Since the polarization state of the light is directly related to the electric field vector, this means that we can also represent the polarization states of the photons by 2-component column vectors or ket vectors of the form

$$|\psi\rangle = \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix} \text{ where we assume the normalization condition } \langle\psi\,|\,\psi\rangle = 1$$

The components in the *state vectors* depend only on the actual polarization state of the photon.

The state vector contains *all* of the information that we can have about the state of polarization of the photon (remember that it consists of just two numbers).

Examples

$$|x\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \text{linear polarized photon}$$

$$|y\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \text{linear polarized photon}$$

$$|R\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \rightarrow \text{right circular polarized photon}$$

$$|L\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} \rightarrow \text{left circular polarized photon}$$

$$|45\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow \text{photon linearly polarized at } 45^\circ \text{ to the x-axis}$$

The bra vector or linear functional corresponding the ket vector $|\psi\rangle$ is given by the *row vector*

$$\langle\psi| = \begin{pmatrix} \psi_x^* & \psi_y^* \end{pmatrix}$$

which clearly implies via our inner product rules

$$\langle\psi\,|\,\psi\rangle = \begin{pmatrix} \psi_x^* & \psi_y^* \end{pmatrix} \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix} = |\psi_x|^2 + |\psi_y|^2 = 1$$

In general, for

$$|\phi\rangle = \begin{pmatrix} \phi_x \\ \phi_y \end{pmatrix}$$

the inner product rule says that

$$\langle\phi\,|\,\psi\rangle = \begin{pmatrix} \phi_x^* & \phi_y^* \end{pmatrix} \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix} = \phi_x^* \psi_x + \phi_y^* \psi_y = \langle\psi\,|\,\phi\rangle^*$$

We also have the results

$$\langle x \,|\, x \rangle = 1 = \langle y \,|\, y \rangle \text{ and } \langle x \,|\, y \rangle = 0 = \langle y \,|\, x \rangle \rightarrow \text{ orthonormal set}$$

$$\langle R \,|\, R \rangle = 1 = \langle L \,|\, L \rangle \text{ and } \langle R \,|\, L \rangle = 0 = \langle L \,|\, R \rangle \rightarrow \text{ orthonormal set}$$

Each of these two sets of state vectors is a basis for the 2-dimensional vector space of polarization states since any other state vector can be written as a linear combination of them, i.e.,

$$|\psi\rangle = \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix} = \psi_x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \psi_y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \psi_x |x\rangle + \psi_y |y\rangle$$

or

$$|\psi\rangle = \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix} = \frac{\psi_x - i\psi_y}{2} \begin{pmatrix} 1 \\ i \end{pmatrix} + \frac{\psi_x + i\psi_y}{2} \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

$$= \frac{\psi_x - i\psi_y}{\sqrt{2}} |R\rangle + \frac{\psi_x + i\psi_y}{\sqrt{2}} |L\rangle$$

We can find the components along the basis vectors using

$$\langle x \,|\, \psi \rangle = \langle x | \left(\psi_x |x\rangle + \psi_y |y\rangle \right) = \psi_x \langle x \,|\, x \rangle + \psi_y \langle x \,|\, y \rangle = \psi_x$$

$$\langle y \,|\, \psi \rangle = \langle y | \left(\psi_x |x\rangle + \psi_y |y\rangle \right) = \psi_x \langle y \,|\, x \rangle + \psi_y \langle y \,|\, y \rangle = \psi_y$$

or

$$|\psi\rangle = |x\rangle \langle x \,|\, \psi \rangle + |y\rangle \langle y \,|\, \psi \rangle$$

and similarly

$$|\psi\rangle = |R\rangle \langle R \,|\, \psi \rangle + |L\rangle \langle L \,|\, \psi \rangle$$

Basically, we are illustrating examples of a superposition principle, which says that any arbitrary polarization state can be written as a superposition (linear combination) of *x*- and *y*-polarization states or equivalently, as a superposition of right- and left-circularly polarized states.

Our earlier discussions of a beam of light passing through a polaroid can now be recast in terms of these polarization states.

Classical physics says that the beam is a superposition of an *x*-polarized beam and a *y*-polarized beam and when this beam passes through an *x*-polaroid, its effect is to remove the *y*-polarized beam and pass the *x*-polarized beam through unchanged.

The energy of the classical beam is given by $|\vec{E}|^2$ which for the polarization states is proportional to $|\psi_x|^2 + |\psi_y|^2$.

Thus, the beam energy after passing through an x-polaroid is proportional to $|\psi_x|^2$. The fraction of the beam energy or the fraction of the number of photons in the beam that passes through is given by

$$\frac{|\psi_x|^2}{|\psi_x|^2 + |\psi_y|^2} = |\psi_x|^2 = |\langle x|\psi\rangle|^2$$

for states normalized to 1.

Our earlier discussion for the case of a single photon forced us to set this quantity equal to the probability of a single photon in the state $|\psi\rangle$ passing through an *x*-polaroid or

$$\begin{aligned}&\text{probability of a photon in the state } |\psi\rangle\\ &\text{passing through an } x\text{-polaroid} = |\langle x|\psi\rangle|^2\end{aligned}$$

This agrees with the mathematical results we derived in earlier lectures.

This all makes sense in an ensemble interpretation of the state vector, that is, that the state vector represents the beam or equivalently, *many copies* of a single photon system.

We then define $\langle x|\psi\rangle$ as the probability amplitude for the individual photon to pass through the x-polaroid.

Another example confirming these results is light passing through a prism.

A prism passes right-circularly-polarized(RCP) light and rejects(absorbs) left-circularly-polarized(LCP) light.

Since we can write

$$|\psi\rangle = |R\rangle\langle R|\psi\rangle + |L\rangle\langle L|\psi\rangle$$

we can generalize the polaroid result to say

$$\begin{aligned}&\text{amplitude that a photon in the state } |\psi\rangle\\ &\text{passes through a prism} = \langle R|\psi\rangle\end{aligned}$$

and

$$\begin{aligned}&\text{probability that a photon in the state } |\psi\rangle\\ &\text{passes through a prism} = |\langle R|\psi\rangle|^2\end{aligned}$$

Polaroids and prisms are examples of *go-nogo* devices.

Certain photons are passed through while others are absorbed in these devices.

We note that the probability is independent of the phase of the state, but the amplitude depends on phase.

This will be a crucial point later on in our discussion.

How Many Basis Sets?

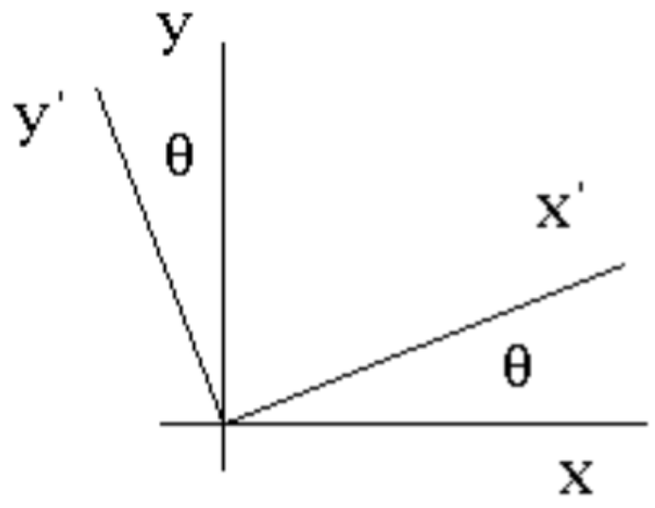
We have already seen two examples of basis sets for the 2-dimensional vector space of polarization states, namely,

$$\{|x\rangle, |y\rangle\} \text{ and } \{|R\rangle, |L\rangle\}$$

In the 2-dimensional vector space there are an infinite number of such orthonormal basis sets related to the $\{|x\rangle,|y\rangle\}$ set.

They are all equivalent for describing physical systems (they correspond to different orientations of the polaroid in the experimental measurement).

We can obtain the other sets say $\{|x'\rangle,|y'\rangle\}$, by a rotation of the bases (or axes) as shown in the Figure below.



Rotation of Axes

We then have in the *x*–*y* basis

$$|\psi\rangle = \psi_x |x\rangle + \psi_y |y\rangle = \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix} = \begin{pmatrix} \langle x|\psi\rangle \\ \langle y|\psi\rangle \end{pmatrix}$$

and if we choose to use the equivalent x′–y′ basis we have

$$|\psi\rangle = \psi_{x'} |x'\rangle + \psi_{y'} |y'\rangle = \begin{pmatrix} \psi_{x'} \\ \psi_{y'} \end{pmatrix} = \begin{pmatrix} \langle x'|\psi\rangle \\ \langle y'|\psi\rangle \end{pmatrix}$$

How are these components related to each other?

We have

$$|\psi\rangle = |x\rangle\langle x|\psi\rangle + |y\rangle\langle y|\psi\rangle$$

which implies

$$\langle x'|\psi\rangle = \langle x'|x\rangle\langle x|\psi\rangle + \langle x'|y\rangle\langle y|\psi\rangle$$

or in matrix notation

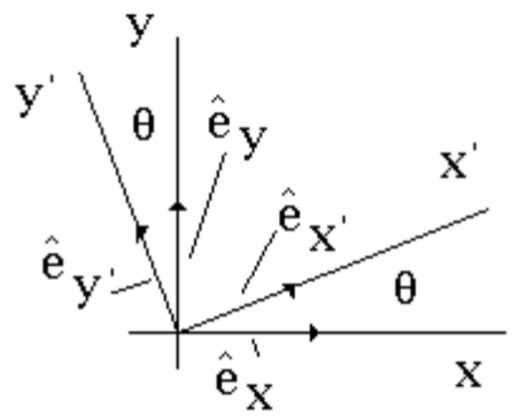
$$\begin{pmatrix}\langle x'|\psi\rangle\\ \langle y'|\psi\rangle\end{pmatrix} = \begin{pmatrix}\langle x'|x\rangle & \langle x'|y\rangle\\ \langle y'|x\rangle & \langle y'|y\rangle\end{pmatrix} = \begin{pmatrix}\langle x|\psi\rangle\\ \langle y|\psi\rangle\end{pmatrix}$$

So we can transform the basis (transform the components) if we can determine the 2×2 transformation matrix

$$\begin{pmatrix}\langle x'|x\rangle & \langle x'|y\rangle\\ \langle y'|x\rangle & \langle y'|y\rangle\end{pmatrix}$$

It turns out that this result is quite general in the sense that it holds for any two bases, not just the linearly polarized bases we used to derive it.

For the linear(plane) polarized case, we can think of an analogy to unit vectors along the axes in ordinary space as shown in the Figure



Rotation of Axes

Then we have(by analogy)

$$\hat{e}_x \cdot \hat{e}_{x'} = \cos\theta = \langle x' \,|\, x \rangle \quad , \quad \hat{e}_{x'} \cdot \hat{e}_y = \sin\theta = \langle x' \,|\, y \rangle$$

$$\hat{e}_{y'} \cdot \hat{e}_y = \cos\theta = \langle y' \,|\, y \rangle \quad , \quad \hat{e}_{y'} \cdot \hat{e}_x = -\sin\theta = \langle y' \,|\, x \rangle$$

or

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

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

and the transformation matrix, $\hat{R}(\theta)$, is

$$\hat{R}(\theta) = \begin{pmatrix} \langle x' \,|\, x \rangle & \langle x' \,|\, y \rangle \\ \langle y' \,|\, x \rangle & \langle y' \,|\, y \rangle \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

so that

$$\begin{pmatrix} \langle x' \,|\, \psi \rangle \\ \langle y' \,|\, \psi \rangle \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \langle x \,|\, \psi \rangle \\ \langle y \,|\, \psi \rangle \end{pmatrix}$$

There are two equivalent ways to interpret these results.

First, we could say it tells us the components of $|\psi\rangle$ in the rotated basis (we keep the vector fixed and rotate the axes).

Second, we can rotate the vector and keep the axes fixed(rotate in the opposite direction).

In this case, we regard

$$\begin{pmatrix} \langle x' \,|\, \psi \rangle \\ \langle y' \,|\, \psi \rangle \end{pmatrix}$$

as a new vector $|\psi'\rangle$ whose components in the fixed x–y basis are the same as the components of $|\psi\rangle$ in the x'–y'-basis or

$$\langle x' \,|\, \psi \rangle = \langle x \,|\, \psi' \rangle \;\text{ and }\; \langle y' \,|\, \psi \rangle = \langle y \,|\, \psi' \rangle$$

For real ψ_x and ψ_y, $|\psi'\rangle$ is the vector $|\psi\rangle$ rotated clockwise by θ or, regarding $\hat{R}(\theta)$ as a linear operator in the vector space we have

$$|\psi'\rangle = \hat{R}(\theta)\,|\psi\rangle$$

It is not a Hermitian operator, however, so it cannot represent an observable.

It is a transformation of vectors and according to our earlier discussion, it should be a unitary operator.

We can see this as follows:

$$\hat{R}^{-1}(\theta) = \hat{R}(-\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \hat{R}^{T}(\theta) = \hat{R}^{\dagger}(\theta)$$

Since $\hat{R}(\theta)$ is a unitary transformation operator for rotations, our earlier discussions say that we should be able to express it as an exponential operator involving the angular momentum with respect to the axis of rotation (z-axis) $\hat{J}_z$, of the form

$$\hat{R}(\theta) = e^{i\theta \hat{J}_z/\hbar}$$

Now, we can rewrite $\hat{R}(\theta)$ as

$$\hat{R}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \cos\theta \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + i\sin\theta \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

$$= \cos\theta \hat{I} + i\sin\theta \hat{Q}$$

where the physical meaning of the operator

$$\hat{Q} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

is yet to be determined.

Expanding the exponential function in a power series we have

$$\hat{R}(\theta) = \hat{R}(0) + \frac{1}{1!}\left.\frac{d\hat{R}(\theta)}{d\theta}\right|_{\theta=0}\theta + \frac{1}{2!}\left.\frac{d^2\hat{R}(\theta)}{d\theta^2}\right|_{\theta=0}\theta^2 + \frac{1}{3!}\left.\frac{d^3\hat{R}(\theta)}{d\theta^3}\right|_{\theta=0}\theta^3 + \frac{1}{4!}\left.\frac{d^4\hat{R}(\theta)}{d\theta^4}\right|_{\theta=0}\theta^4 + \ldots$$

$$= \hat{I} + \frac{1}{1!}\left(\frac{i\hat{J}_z}{\hbar}\right)\theta + \frac{1}{2!}\left(\frac{i\hat{J}_z}{\hbar}\right)^2\theta^2 + \frac{1}{3!}\left(\frac{i\hat{J}_z}{\hbar}\right)^3\theta^3 + \frac{1}{4!}\left(\frac{i\hat{J}_z}{\hbar}\right)^4\theta^4 + \ldots$$

$$= e^{i\theta\hat{J}_z/\hbar}$$

Now let us assume that $\hbar\hat{Q} = \hat{J}_z$, where $\hat{J}_z$ is defined as above.

We then have the following algebraic results.

$$\hat{J}_z^2 = \hbar^2\hat{Q}^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \hbar^2\hat{I}$$

which implies that

$$\hat{J}_z^3 = \hbar^2\hat{J}_z \quad , \quad \hat{J}_z^4 = \hbar^2\hat{J}_z^2 = \hbar^4\hat{I} \quad , \quad \hat{J}_z^5 = \hbar^4\hat{J}_z \text{ and so on}$$

This gives

$$\hat{R}(\theta) = e^{i\theta\hat{J}_z/\hbar} = \hat{I} + (i)\frac{\hat{j}_z}{\hbar}\theta + \frac{(i)^2\hat{I}}{2!}\theta^2 + (i)\frac{\hat{j}_z}{3!\hbar}\theta^3 + \frac{(i)^4\hat{I}}{4!}\theta^4 + \ldots$$

$$= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots\right) + i\left(\theta - \frac{\theta^3}{3!} + \ldots\right)$$

$$= \cos\theta\hat{I} + i\sin\theta\hbar\hat{J}_z = \cos\theta\hat{I} + i\sin\theta\hat{Q}$$

as earlier.

So everything, exponential form of the rotation operator and matrix form of the rotation operator, is consistent if we choose

$$\hat{J}_z = \hbar\hat{Q} = \hbar\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

This corresponds to the matrix representation of $\hat{J}_z$ in the$\{|x\rangle,|y\rangle\}$ basis.

We now work out the eigenvectors and eigenvalues of $\hat{R}(\theta)$ as given by the equation

$$\hat{R}(\theta)\,|\psi\rangle = c\,|\psi\rangle$$

where c = the eigenvalue corresponding to the eigenvector $|\psi\rangle$.

Since all vectors are eigenvectors of the identity operator $\hat{I}$, we only need to find the eigenvectors and eigenvalues of $\hat{J}_z$ in order to solve the problem for $\hat{R}(\theta)$, i.e.,

$$\hat{R}(\theta)\,|\psi\rangle = \cos\theta\hat{I}\,|\psi\rangle + i\sin\theta\hat{Q}\,|\psi\rangle = cos\theta\,|\psi\rangle + \frac{i}{\hbar}\sin\theta\hat{J}_z\,|\psi\rangle = c\,|\psi\rangle$$

If we let(since $\hat{R}(\theta)$ and $\hat{J}_z$ commute they have common eigenvectors)

$$\hat{J}_z\,|\psi\rangle = \lambda\,|\psi\rangle$$

then we have

$$\hat{R}(\theta)\,|\psi\rangle = \cos\theta\hat{I}\,|\psi\rangle + i\sin\theta\hat{Q}\,|\psi\rangle = \cos\theta\,|\psi\rangle + \frac{i}{\hbar}\sin\theta\hat{J}_z\,|\psi\rangle = c\,|\psi\rangle$$

or

$$c = \cos\theta + \frac{i\lambda}{\hbar}\sin\theta$$

Now, since $\hat{J}_z^2 = \hbar^2\hat{I}$ (check it yourself), we have

$$\hat{J}_z^2\,|\psi\rangle = \lambda^2\,|\psi\rangle = \hbar^2\hat{I}\,|\psi\rangle = \hbar^2\,|\psi\rangle$$

which says that

$$\lambda^2 = \hbar^2 \text{ or } \lambda = \pm\hbar = \text{ eigenvalues of } \hat{J}_z$$

We can find the corresponding eigenvectors by inserting the eigenvalues into the eigenvalue equation

$$\hat{J}_z \left| J_z = \hbar \right\rangle = \hbar \left| J_z = \hbar \right\rangle$$

We assume that

$$\left| J_z = \hbar \right\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \text{ where } |a|^2 + |b|^2 = 1$$

to get

$$\hbar \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \hbar \begin{pmatrix} -ib \\ ia \end{pmatrix} = \hbar \begin{pmatrix} a \\ b \end{pmatrix}$$

This gives the result $ia = b$ which together with the normalization condition says that $a = 1/\sqrt{2}$.

We have arbitrarily chosen a to be real since only the relative phase between components is important.

This then gives $b = i/\sqrt{2}$.

Finally, we have the eigenvector

$$\left| J_z = \hbar \right\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} = |R\rangle$$

Similarly, we get

$$\left| J_z = -\hbar \right\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} = |L\rangle$$

So the eigenvectors of $\hat{J}_z$ and hence of $\hat{R}(\theta)$ are the RCP and LCP basis states.

We then have

$$\begin{aligned}\hat{R}(\theta)\,|R\rangle &= \left(\cos\theta\hat{I} + \frac{i}{\hbar}\sin\theta\hat{J}_z\right)|R\rangle \\ &= (\cos\theta + i\sin\theta)\,|R\rangle \\ &= e^{i\theta}\,|R\rangle\end{aligned}$$

Similarly,

$$\hat{R}(\theta)\,|L\rangle = e^{-i\theta}\,|L\rangle$$

This agrees with our earlier discussion where we found that the eigenvalues of a unitary operator have complex exponential form.

Physically, this says that $|R\rangle$ and $|L\rangle$ are only changed by an overall phase factor under a rotation of the basis.

This allows us to easily specify what happens to an arbitrary vector $|\psi\rangle$ under rotations.

The procedure we will use next is, as we will see over and over again, the standard way to do things in quantum mechanics.

First, we expand the arbitrary vector in the $\{|R\rangle,|L\rangle\}$ basis

$$|\psi\rangle = |R\rangle\langle R|\psi\rangle + |L\rangle\langle L|\psi\rangle$$

We then apply the rotation operator to obtain

$$\begin{aligned}\hat{R}(\theta)\,|\psi\rangle &= \hat{R}(\theta)\,|R\rangle\langle R|\psi\rangle + \hat{R}(\theta)\,|L\rangle\langle L|\psi\rangle \\ &= e^{i\theta}\,|R\rangle\langle R|\psi\rangle + e^{-i\theta}\,|L\rangle\langle L|\psi\rangle\end{aligned}$$

or the RCP component of the vector is multiplied by the phase factor $e^{i\theta}$ and the LCP component of the vector is multiplied by a different phase factor $e^{-i\theta}$.

Thus, rotations change the relative phase of the components, which is a real physical change (as opposed to an overall phase change of the state vector).

We interpret (reasons will be clear later) these results to say that the RCP photon is in a state which is an eigenvector of $\hat{J}_z$ with eigenvalue $+\hbar$ or that the photon in that state has z-component of spin = $+\hbar$.

Similarly, a LCP photon has z-component of spin = $-\hbar$.

Now, it is an experimental fact that if a photon traveling in the z-direction is absorbed by matter, then the z-component of the angular momentum of the absorber increases by $\hbar$ or decreases by $\hbar$.

It never remains the same, nor does it change by any value other than $\pm\hbar$.

One cannot predict, for any single photon, whether the change will be $+\hbar$ or $-\hbar$.

We can, however, predict the probability of either value occurring.

In particular, according to our probability formalism, we must have

$$|\langle R|\psi\rangle|^2 = \text{ probability of } +\hbar \quad , \quad |\langle L|\psi\rangle|^2 = \text{ probability of } -\hbar$$

and the average value of the z-component of the angular momentum is

$$\langle \hat{J}_z \rangle = \sum_{all\ possibilities} (\text{eigenvalue}) \times (\text{probability of the eigenvalue})$$

or

$$\langle \hat{J}_z \rangle = +\hbar |\langle R|\psi\rangle|^2 - \hbar |\langle L|\psi\rangle|^2$$

In general, a photon is neither pure RCP nor pure LCP and the angular mo- mentum does not have a definite value.

We can still talk in terms of probabilities, however.

The discreteness of the angular momentum spectrum forces a probabilistic interpretation on us.

We can easily see how all of this works using our mathematical formalism for average values as follows:

$$\begin{aligned}
\langle \hat{J}_z \rangle &= \langle \psi | \hat{J}_z | \psi \rangle \\
&= (\langle \psi | R \rangle \langle R | + \langle \psi | L \rangle \langle L |) \hat{J}_z (| R \rangle \langle R | \psi \rangle + | L \rangle \langle L | \psi \rangle) \\
&= (\langle R | \psi \rangle^* \langle R | + \langle L | \psi \rangle^* \langle L |) \hat{J}_z (| R \rangle \langle R | \psi \rangle + | L \rangle \langle L | \psi \rangle) \\
&= \langle R | \hat{J}_z | R \rangle | \langle R | \psi \rangle |^2 + \langle L | \hat{J}_z | L \rangle | \langle L | \psi \rangle |^2 \\
&\quad + \langle R | \hat{J}_z | L \rangle \langle R | \psi \rangle^* \langle L | \psi \rangle + \langle L | \hat{J}_z | R \rangle \langle L | \psi \rangle^* \langle R | \psi \rangle \\
&= +\hbar | \langle R | \psi \rangle |^2 - \hbar | \langle L | \psi \rangle |^2
\end{aligned}$$

as we stated earlier.

Let us return for a moment to the matrix representation of the $\hat{J}_z$ operator.

We have found the following results:

$$\hat{J}_z | R \rangle = +\hbar | R \rangle \text{ and } \hat{J}_z | L \rangle = -\hbar | L \rangle$$

In the $\{|R\rangle, |L\rangle\}$ basis, these relations imply the matrix representation

$$\hat{J}_z = \begin{pmatrix} \langle R | \hat{J}_z | R \rangle & \langle R | \hat{J}_z | L \rangle \\ \langle L | \hat{J}_z | R \rangle & \langle L | \hat{J}_z | L \rangle \end{pmatrix} = \hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

which is the standard form of $\hat{J}_z$ in terms of one of the so-called *Pauli matrices*, namely,

$$\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \rightarrow \hat{J}_z = \hbar\hat{\sigma}_z$$

Now

$$|x\rangle = \frac{1}{\sqrt{2}}(|R\rangle + |L\rangle) \text{ and } |y\rangle = \frac{i}{\sqrt{2}}(|R\rangle - |L\rangle)$$

and, therefore, in the $\{|x\rangle,|y\rangle\}$ basis we have the matrix representation

$$\hat{J}_z = \begin{pmatrix} \langle x|\hat{J}_z|x\rangle & \langle x|\hat{J}_z|y\rangle \\ \langle y|\hat{J}_z|x\rangle & \langle y|\hat{J}_z|y\rangle \end{pmatrix} = \hbar\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

which is the form we *guessed and used* earlier.

Projection Operators

Let us now turn our attention to projection operators and density operators in the context of photon polarization.

The general operator $|\psi\rangle\langle\phi|$ can be represented by a 2×2 matrix in the polar ization state vector space.

It is constructed using the *outer product rule*:

$$\hat{P} = |\psi\rangle\langle\phi| = \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix}\begin{pmatrix} \phi_x^* & \phi_y^* \end{pmatrix} = \begin{pmatrix} \psi_x\phi_x^* & \psi_x\phi_y^* \\ \psi_y\phi_x^* & \psi_y\phi_y^* \end{pmatrix}$$

or equivalently, by choosing a basis and finding the matrix representation

$$\hat{P} = \begin{pmatrix} \langle x|\hat{P}|x\rangle & \langle x|\hat{P}|y\rangle \\ \langle y|\hat{P}|x\rangle & \langle y|\hat{P}|y\rangle \end{pmatrix} = \begin{pmatrix} \langle x|\psi\rangle\langle\phi|x\rangle & \langle x|\psi|\langle\phi|y\rangle \\ \langle y|\psi||\phi|x\rangle & \langle y|\psi||\phi|y\rangle \end{pmatrix}$$

$$= \begin{pmatrix} \psi_x\phi_x^* & \psi_x\phi_y^* \\ \psi_y\phi_x^* & \psi_y\phi_y^* \end{pmatrix}$$

In particular, we have for the projection operators

$$|x\rangle\langle x| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad , \quad |x\rangle\langle y| = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$|y\rangle\langle x| = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad , \quad |y\rangle\langle y| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

From these results we easily see that

$$|x\rangle\langle x| + |y\rangle\langle y| = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \hat{I}$$

and

$$|\psi\rangle = \hat{I}\,|\psi\rangle = (|x\rangle\langle x| + |y\rangle\langle y|)\,|\psi\rangle$$

$$= |x\rangle\langle x\,|\,\psi\rangle + |y\rangle\langle y\,|\,\psi\rangle = \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix}$$

as they should.

Similarly, we have

$$|R\rangle\langle R| + |L\rangle\langle L| = \hat{I}$$

which leads to

$$\hat{J}_z = \hat{J}_z\hat{I} = \hat{J}_z(|R\rangle\langle R| + |L\rangle\langle L|)$$

$$= \hbar\,|R\rangle\langle R| - \hbar\,|L\rangle\langle L|$$

which is the expansion of the operator $\hat{J}_z$ in terms of eigenvalues and 1-dimensional subspace projection operators (eigenvectors) that we discussed earlier.

Amplitudes and Probabilities

The *probability interpretation* we have been making follows from the concept of *superposition*.

The superposition idea says that we can write any arbitrary photon state as a linear combination of basis states

$$|\psi\rangle = |R\rangle\langle R|\psi\rangle + |L\rangle\langle L|\psi\rangle$$

We then interpreted $|\langle R|\psi\rangle|^2$ as the probability that the photon in the state $|\psi\rangle$ will behave as a RCP photon in the state $|R\rangle$.

Generalizing this statement, we say that a system in a state $|\psi\rangle$, in Quantum Mechanics, has a probability $|\langle\phi|\psi\rangle|^2$ of behaving like it was in the state $|\phi\rangle$.

You might now conclude, from the experimental fact that only $\pm\hbar$ is transferred to matter by light, that photons are *always* either in the state $|R\rangle$ with some probability α or in the state $|L\rangle$ with probability $1-\alpha$.

FACT: An x-polarized photon never passes through a y-polaroid

PROBLEM: If, the above interpretation of being either $|R\rangle$ or $|L\rangle$ was true, then

1. an x-polarized photon has a probability $|\langle R|x\rangle|^2 = 1/2$ of being RCP and a RCP photon has a probability $|\langle y|R\rangle|^2 = 1/2$ of being a y-polarized photon and thus passing through a y-polaroid.

2. an x-polarized photon has a probability $|\langle L|x\rangle|^2 = 1/2$ of being LCP and a LCP photon has a probability $|\langle y|L\rangle|^2 = 1/2$ of being a y-polarized photon and thus passing through a y-polaroid.

This means that if we assume we can think that the photon is either $|R\rangle$ or $|L\rangle$ but we do not know which, i.e., photon properties have an objective reality, then the total probability that an x-polarized photon would get through a y-polaroid in this interpretation is

$$\text{total probability} = |\langle R|x\rangle|^2 |\langle y|R\rangle|^2 + |\langle L|x\rangle|^2 |\langle y|L\rangle|^2 = \frac{1}{2}$$

However, as we stated, it **NEVER HAPPENS**.

What is wrong?

SOLUTION: When we think of an x-polarized photon as being a RCP photon or a LCP photon with equal probability, we are ruling out the possibility of any *interference* between the RCP and LCP amplitudes.

We are thinking classically!

We give meaning to the word *interference* here in this way.

The correct calculation of the probability, which lays the foundation for all of the amplitude mechanics rules in Quantum Mechanics, goes as follows:

1. The probability amplitude of an x-polarized photon passing through a y-polaroid $= \langle y|x\rangle = 0$, which implies that the probability $|\langle y|x\rangle|^2 = 0$ also.

2. If we say that the x-polarized photon is in a *superposition* of $|R\rangle$ and $|L\rangle$ (we make no statement about probabilities at this point), this implies that

$$|x\rangle = |R\rangle\langle R|x\rangle + |L\rangle\langle L|x\rangle \tag{5.89}$$

which gives

$$\langle y|x\rangle = \langle y|R\rangle\langle R|x\rangle + \langle y|L\rangle\langle L|x\rangle \tag{5.90}$$

or the *amplitude* for an x-polarized photon to pass through a y-polaroid is the *sum of two amplitudes*, namely, that it passes through as a RCP photon $\langle y|R\rangle\langle R|x\rangle$ and that it passes through as a LCP photon $\langle y|L\rangle\langle L|x\rangle$.

3. The *probability* of passing through is then the *absolute square of the total amplitude*

$$\begin{aligned}
\textit{probability} &= |\langle y|R\rangle\langle R|x\rangle + \langle y|L\rangle\langle L|x\rangle|^2 \\
&= (\langle y|R\rangle^*\langle R|x\rangle^* + \langle y|L\rangle^*\langle L|x\rangle^*)(\langle y|R\rangle\langle R|x\rangle + \langle y|L\rangle\langle L|x\rangle) \\
&= |\langle R|x\rangle|^2|\langle y|R\rangle|^2 + |\langle L|x\rangle|^2|\langle y|L\rangle|^2 \\
&\quad \langle y|R\rangle\langle R|x\rangle\langle y|L\rangle^*\langle L|x\rangle^* + \langle y|R\rangle^*\langle R|x\rangle^*\langle y|L\rangle\langle L|x\rangle
\end{aligned}$$

4. The first two terms are the same as the incorrect calculation (6.88) above. The last two terms represent *interference effects* between the two amplitudes (RCP way and LCP way).

A simple calculation shows that the interference terms exactly cancel the first two terms and that the probability equals zero in agreement with experiment!

INTERPRETATION: The way to interpret this result is as follows:

$$\langle y|R\rangle\langle R|x\rangle = \text{probability amplitude for an x-polarized photon to pass through a y-polaroid as a RCP photon}$$

$$\langle y|L\rangle\langle L|x\rangle = \text{probability amplitude for an x-polarized photon to pass through a y-polaroid as a LCP photon}$$

These are *indistinguishable* ways for the process to occur, i.e., no measurement exists that can tell us whether it passes through the system as an RCP photon or as a LCP photon without destroying the interference, i.e., without radically altering the experiment.

To get the correct total probability, we *add all the amplitudes for indistinguishable ways* and then square the resulting total amplitude.

In the incorrect calculation, we found the probability for each indistinguishable way and then added the probabilities.

In one case, we eliminated the interference effects and got the wrong result and, in the other case, we included the interference effects and obtained the correct result.

Summarizing, we have these rules for amplitude mechanics and probabilities in Quantum Mechanics:

1. The probability amplitude for two successive events is the *product* of the amplitudes for each event, i.e., the amplitude for the x-polarized photon to pass through the y-polaroid as a RCP polarized photon is the product of the amplitude for an x-polarized photon to be a RCP photon $\langle R|x\rangle$ and the amplitude for a RCP photon to be a y-polarized photon $\langle y|R\rangle$

$$\langle R|x\rangle\langle y|R\rangle$$

2. The total amplitude for a process that can take place in several *indistinguishable* ways is the *sum* of the amplitudes for each individual way, i.e.,

$$\langle y|x\rangle = \langle R|x\rangle\langle y|R\rangle + \langle L|x\rangle\langle y|L\rangle$$

We note here that this is merely a reflection of the property of projection operators that

$$\hat{I} = |R\rangle\langle R| + |L\rangle\langle L|$$

which says that

$$\langle y|x\rangle = \langle y|\hat{I}|x\rangle = \langle R|x\rangle\langle y|R\rangle + \langle L|x\rangle\langle y|L\rangle$$

Thus, the mathematical sum over all projection operators being equal to the identity operator is physically equivalent to the sum over all possible intermediate states and it turns into a sum over all the amplitudes for indistinguishable ways in this interpretation.

3. The total probability for the process to occur is the absolute square of the total amplitude.

So, in classical physics, we

1. find amplitudes and probabilities of each way separately
2. add all probabilities to get total probability

We get **NO** interference effects!!

In Quantum Mechanics, we

1. find the amplitudes for each indistinguishable way the process can occur
2. add all the amplitudes to get a total amplitude
3. square the total amplitude to get the total probability

The important result here is that we must consider **ALL INDISTINGUISHABLE WAYS** in step (2).

An *indistinguishable* way is characterized as follows:

1. If two ways are indistinguishable, then there exists *no measurement* that can decide which of the two ways actually happened without altering the experiment.
2. In particular, if we attempt to find out, then the interference effects will disappear and we will return to the classical result obtained by adding probabilities.

What actually happens is that during any measurement trying distinguish the ways, the relative phase of the components in the superposition becomes completely uncertain and this will wash out the interference.

This happens as follows: instead of

$$|x\rangle = |R\rangle\langle R|x\rangle + |L\rangle\langle L|x\rangle$$

we would have, if we attempted to add a measurement to determine if the *x*-polarized photon was RCP or LCP, to consider the state

$$|\tilde{x}\rangle = e^{i\alpha_R}|R\rangle\langle R|x\rangle + e^{i\alpha_L}|L\rangle\langle L|x\rangle$$

A probability calculation then gives

$$\begin{aligned} total\ probability\ &= |\langle y|R\rangle\langle R|x\rangle + \langle y|L\rangle\langle L|x\rangle|^2 \\ &\quad + 2Real\left[\langle y|R\rangle\langle R|x\rangle e^{i(\alpha_R-\alpha_L)}\langle y|L\rangle^*\langle L|x\rangle^*\right] \end{aligned}$$

The observed probability, which is the result of many *identical* measurements in the laboratory(in the standard interpretation), is an average over all values of the extra phases(they are random in different experiments).

This involves integrating over the relative phase, i.e.,

$$\frac{1}{2\pi}\int_0^{2\pi} e^{i(\alpha_R-\alpha_L)}\,d(\alpha_R-\alpha_L) = 0$$

It is clear that the interference term averages to zero and we get the classical result!

This means we *cannot add* such a measurement and retain any quantum results!

Pure States, Unpure States and Density Operators

If the photon were in the state $|x\rangle$, then we would have, for some linear operator $\hat{A}$

$$\langle \hat{A} \rangle = \langle x | \hat{A} | x \rangle$$

From our earlier discussion of density operators, however, we must also have,

$$\langle \hat{A} \rangle = Tr(\hat{W}\hat{A})$$

where $\hat{W}$ is a density operator.

Therefore we must have (using the $\{|x\rangle,|y\rangle\}$ basis)

$$\begin{aligned}\langle \hat{A} \rangle = \langle x | \hat{A} | x \rangle = Tr(\hat{W}\hat{A}) &= \langle x | \hat{W}\hat{A} | x \rangle + \langle y | \hat{W}\hat{A} | y \rangle \\ &= \langle x | \hat{W}\hat{I}\hat{A} | x \rangle + \langle y | \hat{W}\hat{I}\hat{A} | y \rangle \\ &= \langle x | \hat{W} | x \rangle \langle x | \hat{A} | x \rangle + \langle x | \hat{W} | y \rangle \langle y | \hat{A} | x \rangle \\ &\quad + \langle y | \hat{W} | x \rangle \langle x | \hat{A} | y \rangle + \langle y | \hat{W} | y \rangle \langle y | \hat{A} | y \rangle\end{aligned}$$

This implies that

$$\langle x | \hat{A} | x \rangle = 1 \text{ and } \langle y | \hat{A} | x \rangle = \langle x | \hat{A} | y \rangle = \langle y | \hat{A} | y \rangle = 0$$

or

$$\hat{W} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = |x\rangle\langle x|$$

which says that $|x\rangle$ is a pure state.

Now suppose that the photon is in the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}|x\rangle + \frac{1}{\sqrt{2}}|y\rangle$$

This says that the probability = 1/2 that the photon behaves like $|x\rangle$ and the probability = 1/2 that it behaves like $|y\rangle$.

Note that the relative phase between the components is assumed to be known exactly in this state.

In this case, we have

$$\begin{aligned}\langle\hat{A}\rangle = \langle\psi|\hat{A}|\psi\rangle &= \frac{1}{2}\left[\langle x|\hat{A}|x\rangle + \langle x|\hat{A}|y\rangle + \langle y|\hat{A}|x\rangle + \langle y|\hat{A}|y\rangle\right] \\ &= Tr(\hat{W}\hat{A}) = \langle x|\hat{W}\hat{A}|x\rangle + \langle y|\hat{W}\hat{A}|y\rangle = \langle x|\hat{W}\hat{I}\hat{A}|x\rangle + \langle y|\hat{W}\hat{I}\hat{A}|y\rangle \\ &= \langle x|W|x\rangle\langle x|A|x\rangle + \langle x|W|y\rangle\langle y|A|x\rangle \\ &\quad + \langle y|\hat{W}|y\rangle\langle x|\hat{A}|x\rangle + \langle y|\hat{W}|y\rangle\langle y|\hat{A}|y\rangle\end{aligned}$$

which implies that

$$\langle x|\hat{W}|x\rangle = \frac{1}{2} = \langle y|\hat{W}|y\rangle = \langle y|\hat{W}|x\rangle = \langle x|\hat{W}|y\rangle$$

or

$$\hat{W} = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1\\ 1\end{pmatrix}\begin{pmatrix}1 & 1\end{pmatrix} = |\psi\rangle\langle\psi|$$

So, again we have a pure state.

But what happens if we *only know* that the probability = 1/2 that the photon behaves like $|x\rangle$ and the probability = 1/2 that it behaves like $|y\rangle$.

This says that we might write the state vector as

$$|\psi\rangle = a\,|x\rangle + b\,|y\rangle$$

where we only know that $|a|^2 = |b|^2 = 1/2$.

Let us choose

$$a = \frac{e^{i\alpha_a}}{\sqrt{2}} \text{ and } b = \frac{e^{i\alpha_b}}{\sqrt{2}}$$

We do not have any phase information in this case.

In addition, the phases values could be different in each separate experiment.

This last fact means that we must average over the relative phase (the only meaningful phase) $\alpha_a - \alpha_b$ when computing the probabilities and this means that all interference effects will vanish.

When we calculate the expectation value we have

$$\begin{aligned}\langle \hat{A} \rangle &= \langle \psi | \hat{A} | \psi \rangle \\ &= \frac{1}{2}\left[\langle x | \hat{A} | x \rangle + \langle x | \hat{A} | y \rangle e^{-i(\alpha_a - \alpha_b)} + \langle y | \hat{A} | x \rangle e^{i(\alpha_a - \alpha_b)} + \langle y | \hat{A} | y \rangle\right]\end{aligned}$$

and when we average over the relative phase we obtain

$$\langle \hat{A} \rangle = \frac{1}{2}\langle x | \hat{A} | x \rangle + \frac{1}{2}\langle y | \hat{A} | y \rangle$$

Again, we must have

$$\langle \hat{A} \rangle = Tr(\hat{W}\hat{A}) = \langle x|\hat{W}\hat{A}|x\rangle + \langle y|\hat{W}\hat{A}|y\rangle = \langle x|\hat{W}\hat{I}\hat{A}|x\rangle + \langle y|\hat{W}\hat{I}\hat{A}|y\rangle$$
$$= \langle x|\hat{W}|x\rangle\langle x|\hat{A}|x\rangle + \langle x|\hat{W}|y\rangle\langle y|\hat{A}|x\rangle$$
$$+ \langle y|\hat{W}|y\rangle\langle x|\hat{A}|x\rangle + \langle y|\hat{W}|y\rangle\langle y|\hat{A}|y\rangle$$

which implies that

$$\langle x|\hat{W}|x\rangle = \frac{1}{2} = \langle y|\hat{W}|y\rangle \text{ and } \langle y|\hat{W}|x\rangle = \langle x|\hat{W}|y\rangle = 0$$

or

$$\hat{W} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{2}|x\rangle\langle x| + \frac{1}{2}|y\rangle\langle y|$$
$$= Prob(x)|x\rangle\langle x| + Prob(y)|y\rangle\langle y|$$

This is a *nonpure* or *mixed state*.

So, *we have a pure state only if the relative phase information is known exactly*.

The way to describe a nonpure state is by the corresponding density matrix, which only requires knowing probabilities and not phases.

It really does not have a state vector.

Unpolarized Light

Consider the following experiment.

We have a beam of monochromatic light that is composed of photons from two sources which output photons in the states $|\psi_1\rangle$ or $|\psi_2\rangle$, respectively.

The sources emit the photons randomly and are independent of each other, which implies that we cannot tell which source any particular photon comes from.

We assign these probabilities

$$p_1 = \text{probability that a photon comes from source \#1}$$
$$p_2 = \text{probability that a photon comes from source \#2}$$

where $p_1 + p_2 = 1$.

Now the probability that a particular observed photon transfers $\hbar$ is

$$p_+ = p_1 |\langle R|\psi_1\rangle|^2 + p_2 |\langle R|\psi_2\rangle|^2$$

and the probability that it transfers $-\hbar$ is

$$p_- = p_1 |\langle L|\psi_1\rangle|^2 + p_2 |\langle L|\psi_2\rangle|^2$$

This implies that the average value of the angular momentum transfer for the beam of photons is

$$\begin{aligned}
\langle \hat{J}_z \rangle &= \hbar p_+ - \hbar p_- \\
&= \hbar p_1 |\langle R|\psi_1\rangle|^2 + \hbar p_2 |\langle R|\psi_2\rangle|^2 - \hbar p_1 |\langle L|\psi_1\rangle|^2 - \hbar p_2 |\langle L|\psi_2\rangle|^2 \\
&= p_1 \left[\hbar |\langle R|\psi_1\rangle|^2 - \hbar |\langle L|\psi_1\rangle|^2\right] + p_2 \left[\hbar |\langle R|\psi_2\rangle|^2 - \hbar |\langle L|\psi_2\rangle|^2\right] \\
&= p_1 \langle \hat{J}_z \rangle_1 + p_2 \langle \hat{J}_z \rangle_2
\end{aligned}$$

or, the average value of the angular momentum transfer for the beam of photons = sum over the average value in each beam weighted by the probability that photon comes from that beam.

Let me emphasize(once again) at this point that it is important to realize that the statement

$$\text{The photon is either in the state } |\psi_1\rangle \text{ or } |\psi_2\rangle \text{ but we do not know which}$$

is **NOT** the same statement as

$$\text{The photon is in a state which is a superposition of } |\psi_1\rangle \text{ and } |\psi_2\rangle$$

In the second case, we are saying the relative phase is known as in the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}|x\rangle + \frac{1}{\sqrt{2}}|y\rangle$$

which we found to be a pure state.

Being in a superposition implies that we know the relative phase of the components.

In the first case, however, we are saying that the relative phase is unknown and, as we have seen, interference effects will vanish.

We can only specify a density matrix in this case.

In pure states, we have superpositions and the probability amplitude rules apply.

In nonpure or mixed states, where the system is in one of several states with definite probabilities, we find weighted averages (weighted with the state probabilities) of the value in each state.

We use addition of probabilities with no interference effects, which as we have seen, is equivalent to saying the relative phase is unknown.

Unpolarized light has equal probability of being in **any** polarization state.

It is just a special nonpure or mixed state.

No relative phase information is known for unpolarized light.

How Does the Polarization State Vector Change?

Up to now we have been considering devices such as polaroids and prisms, which are *go-nogo* devices.

Some photons get through and some do not for each of these devices depending on their polarization state.

We now consider devices where all the photons get through no matter what their polarization state is, but, during transit, the device changes the incident polarization state in some way.

In particular, we consider the example of a *birefringent* crystal, such as calcite.

A calcite crystal has a *preferred* direction called the *optic* axis.

The crystal has a different index of refraction for light polarized parallel to the optic axis than it has for light polarized perpendicular to the optic axis.

We assume that the optic axis is in the x-y plane and send a beam of photons in the z-direction.

Photons polarized perpendicular to the optic axis are called *ordinary* and are in the state $|o\rangle$ and photons polarized parallel to the optic axis are called *extraordinary* and are in the state $|e\rangle$.

The set $\{|o\rangle,|e\rangle\}$ forms an orthonormal basis and general photon states interacting with a calcite crystal are written as superpositions of these basis states.

The next discussion is an example of a general rule in quantum mechanics.

If we are doing an experiment using a particular measuring device that measures the observable $\hat{Q}$, then we should use as the basis for all states, the eigenvectors of $\hat{Q}$.

As we shall see, this requirement pushes us to ask the correct experimental questions (those that quantum mechanics can answer).

This particular basis is called the *home* space for the experiment.

Now the phase of a light wave with wavelength λ as it propagates through a medium in the z-direction is given by the quantity

$$\phi = e^{ikz}$$

with

$$k = \frac{2\pi}{\lambda} = \frac{n\omega}{c}$$

where n = index of refraction, $\omega = 2\pi f$, f = frequency and c = speed of light.

Since the phase depends on the index of refraction, the effect of passing through a calcite crystal is to change the *relative phase* of the $|o\rangle$ and $|e\rangle$ components making up the superposition, which is a real physical change that is measurable.

We assume that the photon entering the calcite crystal is in the initial state

$$|\psi_{in}\rangle = |e\rangle\langle e|\psi_{in}\rangle + |o\rangle\langle o|\psi_{in}\rangle$$

The two components have different indices of refraction n_e and n_o, respectively.

If the beam passes through a length ℓ then the state upon leaving is given by (remember the component phases change differently)

$$|\psi_{out}\rangle = e^{ik_e\ell}|e\rangle\langle e|\psi_{in}\rangle + e^{ik_o\ell}|o\rangle\langle o|\psi_{in}\rangle = \hat{U}_\ell|\psi_{in}\rangle$$

where

$$\hat{U}_z = e^{ik_e z} |e\rangle \langle e| + e^{ik_o z} |o\rangle \langle o|$$

is a *time development* operator of some sort since ℓ = distance traveled in a time t is proportional to t.

Now we define two new quantities which will be with us throughout our study of Quantum Mechanics.

For transitions between two states (in and out in this case)

$$\langle \phi | \psi_{out} \rangle = \langle \phi | \hat{U}_z | \psi_{in} \rangle = \text{the transition amplitude for a photon to enter the calcite in state } |\psi_{in}\rangle \text{ and leave in state } |\phi\rangle$$

$$| \langle \phi | \psi_{out} |^2 \rangle = | \langle \phi | \hat{U}_z | \psi_{in} \rangle |^2 = \text{the corresponding transition probability}$$

To proceed any further, we need to find out more about the operator $\hat{U}_z$.

Now

$$|\psi_z\rangle = \text{state of photon after traveling a distance z through calcite}$$

$$= \hat{U}_z |\psi_{in}\rangle$$

From the form of $\hat{U}_z$ we have

$$\hat{U}_{z+\epsilon} = e^{ik_e(z+\epsilon)} |e\rangle \langle e| + e^{ik_o(z+\epsilon)} |o\rangle \langle o|$$

$$= (e^{ik_e\epsilon} |e\rangle \langle e| + e^{ik_o\epsilon} |o\rangle \langle o|)(e^{ik_e z} |e\rangle \langle e| + e^{ik_o z} |o\rangle \langle o|)$$

or

$$\hat{U}_{z+\epsilon} = \hat{U}_\epsilon \hat{U}_z$$

This implies that

$$|\psi_{z+\epsilon}\rangle = \hat{U}_{z+\epsilon} |\psi_{in}\rangle = \hat{U}_\epsilon \hat{U}_z |\psi_{in}\rangle = \hat{U}_\epsilon |\psi_z\rangle$$

Now let $\varepsilon \to 0$ such that $k_o\varepsilon \ll 1$ and $k_e\varepsilon \ll 1$ and we can then write (to 1st-order)

$$\begin{aligned}\hat{U}_\epsilon &= e^{ik_e\epsilon}\,|e\rangle\langle e| + e^{ik_o\epsilon}\,|o\rangle\langle o| \\ &= (1+ik_e\epsilon)\,|e\rangle\langle e| + (1+ik_o\epsilon)\,|o\rangle\langle o| \\ &= \hat{I} + i\epsilon\hat{K}\end{aligned}$$

where

$$\hat{I} = |e\rangle\langle e| + |o\rangle\langle o| \text{ and } \hat{K} = k_e\,|e\rangle\langle e| + k_o\,|o\rangle\langle o|$$

Now, the relation

$$\hat{K} = k_e\,|e\rangle\langle e| + k_o\,|o\rangle\langle o|$$

is an expansion of an operator in terms of its eigenvalues and the corresponding projection operators (eigenvectors).

It says that the eigenvectors of $\hat{K}$ are $|e\rangle$ and $|o\rangle$ with eigenvalues k_e and k_o, respectively.

This illustrates the awesome power in these methods we have developed!!

We then have

$$|\psi_{z+\epsilon}\rangle = (\hat{I} + i\epsilon\hat{K})\,|\psi_z\rangle$$

or

$$|\psi_{z+\epsilon}\rangle - |\psi_z\rangle = i\epsilon\hat{K}\,|\psi_z\rangle$$

or

$$\lim_{\epsilon\to 0}\frac{|\psi_{z+\epsilon}\rangle - |\psi_z\rangle}{\epsilon} = i\hat{K}\,|\psi_z\rangle$$

which gives the differential equation for the *time* development of the state vector

$$\frac{d}{dt}\,|\psi_z\rangle = i\hat{K}\,|\psi_z\rangle$$

It is clearly similar to the differential equation we obtained earlier from the time development operator.

If we follow the results from the earlier case, then we should have

$$\hat{K} = \text{Hermitian operator and } \hat{U}_z = \text{unitary operator}$$

Let us derive these two results.

We have, using the $x - y$ basis

$$\begin{aligned}\langle x\,|\,\psi_{z+\epsilon}\rangle - \langle x\,|\,\psi_z\rangle &= i\epsilon\,\langle x|\,\hat{K}\,|\psi_z\rangle \\ &= i\epsilon\,\langle x|\,\hat{K}\hat{I}\,|\psi_z\rangle = i\epsilon\,\langle x|\,\hat{K}\,|x\rangle\,\langle x\,|\,\psi_z\rangle + i\epsilon\,\langle x|\,\hat{K}\,|y\rangle\,\langle y\,|\,\psi_z\rangle\end{aligned}$$

or the change in the x-component of ψ_z as we move an infinitesimal amount ε has one part proportional to the x-component of ψ_z and a second part proportional to the y-component of ψ_z.

Similarly, we have

$$\begin{aligned}\langle y\,|\,\psi_{z+\epsilon}\rangle - \langle y\,|\,\psi_z\rangle &= i\epsilon\,\langle y|\,\hat{K}\,|\psi_z\rangle \\ &= i\epsilon\,\langle y|\,\hat{K}\hat{I}\,|\psi_z\rangle = i\epsilon\,\langle y|\,\hat{K}\,|x\rangle\,\langle x\,|\,\psi_z\rangle + i\epsilon\,\langle y|\,\hat{K}\,|y\rangle\,\langle y\,|\,\psi_z\rangle\end{aligned}$$

Now, since *no photons are lost* as we pass through, we must have

$$\langle\psi_{z+\epsilon}\,|\,\psi_{z+\epsilon}\rangle = \langle\psi_z\,|\,\psi_z\rangle = 1$$

for all z.

We then get

$$\begin{aligned}\langle\psi_{z+\epsilon}\,|\,\psi_{z+\epsilon}\rangle = \langle\psi_z\,|\,\psi_z\rangle &+ i\epsilon\left[\langle x|\,\hat{K}\,|x\rangle - \langle x|\,\hat{K}\,|x\rangle^*\right]|\,\langle x\,|\,\psi_z\rangle^2 \\ &+ i\epsilon\left[\langle y|\,\hat{K}\,|y\rangle - \langle y|\,\hat{K}\,|y\rangle^*\right]|\,\langle y\,|\,\psi_z\rangle^2 \\ &+ i\epsilon\left[\langle x|\,\hat{K}\,|y\rangle - \langle x|\,\hat{K}\,|y\rangle^*\right]\langle y\,|\,\psi_z\rangle\,\langle x\,|\,\psi_z\rangle^* \\ &+ i\epsilon\left[\langle y|\,\hat{K}\,|x\rangle - \langle y|\,\hat{K}\,|x\rangle^*\right]\langle x\,|\,\psi_z\rangle\,\langle y\,|\,\psi_z\rangle^*\end{aligned}$$

which says that we must have

$$\langle x|\hat{K}|x\rangle = \langle x|\hat{K}|x\rangle^* \quad , \quad \langle y|\hat{K}|y\rangle = \langle y|\hat{K}|y\rangle^*$$
$$\langle x|\hat{K}|y\rangle = \langle x|\hat{K}|y\rangle^* \quad , \quad \langle y|\hat{K}|x\rangle = \langle y|\hat{K}|x\rangle^*$$

or that $\hat{K}$ is Hermitian.

Finally, one can show that $\hat{U}_z^\dagger \hat{U}_z = \hat{I}$ so that $\hat{U}_z$ is unitary as expected.

From our earlier discussions, we then identify

$$\hat{U}_z = \text{transformation operator and } \hat{K} = \text{generator of transformation}$$

Calculating the Transition Probability

We defined the *transition probability* as

$$T(z) = |\langle\phi|\psi_{z,out}\rangle|^2 = |\langle\phi|\hat{U}_z|\psi_{z,in}\rangle|^2$$

Using

$$\hat{U}_z = e^{ik_e z}|e\rangle\langle e| + e^{ik_o z}|o\rangle\langle o|$$

and

$$|\psi_{in}\rangle = a|o\rangle + b|e\rangle \text{ where } |a|^2 + |b|^2 = 1$$

we get

$$\begin{aligned} T(z) &= |\langle\phi|(e^{ik_e z}|e\rangle\langle e| + e^{ik_o z}|o\rangle\langle o|)(a|o\rangle + b|e\rangle)|^2 \\ &= |\langle\phi|(be^{ik_e z}|e\rangle + ae^{ik_o z}|o\rangle|^2 \\ &= |be^{ik_e z}\langle\phi|e\rangle + ae^{ik_o z}\langle\phi|o\rangle|^2 \end{aligned}$$

Now let us ask a specific question.

Suppose $a = 1/\sqrt{2} = -ib$, which means the that photon entering the calcite crystal is an LCP photon.

What is the probability that it will exit as a RCP photon?

This means we choose

$$|\phi\rangle = \frac{1}{\sqrt{2}}(|o\rangle + i\,|e\rangle)$$

or

$$\langle\phi\,|\,e\rangle = -\frac{i}{\sqrt{2}} \text{ and } \langle\phi\,|\,o\rangle = \frac{1}{\sqrt{2}}$$

We then get

$$T(z) = |be^{ik_e z}\langle\phi\,|\,e\rangle + ae^{ik_o z}\langle\phi\,|\,o\rangle|^2$$

$$= |\frac{i}{\sqrt{2}}e^{ik_e z}\frac{i}{\sqrt{2}} + \frac{1}{\sqrt{2}}e^{ik_o z}\frac{1}{\sqrt{2}}|^2 = \frac{1}{4}|e^{ik_o z} - -e^{ik_e z}|^2$$

$$= \frac{1}{4}(1 + 1 - e^{i(k_o - k_e)z} - e^{-i(k_o - k_e)z})$$

$$= \frac{1}{2}(1 - \cos(k_o - k_e)z)$$

If we choose $(k_o - k_e)z = \pi$, then $T = 1$ and all the LCP photons are turned into RCP photons by a calcite crystal of just the right length.

This simple example clearly exhibits the power of these techniques

The Strange World of Neutral K-Mesons

We can now use the same formalism we developed for photon polarization to study elementary particles called K-mesons.

K-mesons are produced in high-energy accelerators via the production process

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

In this reaction, electric charge is conserved.

This reaction takes place via the so-called strong interactions.

Another physical quantity called strangeness is also conserved in strong interactions.

All K^0-mesons have a strangeness equal to +1.

For every particle there always exists an antiparticle.

For the K^0, the antiparticle is called the $\bar{K}^0$.

The $\bar{K}^0$-mesons have a strangeness equal to -1.

A reaction involving the $\bar{K}^0$ is

$$\bar{K}^0 + p^+ \rightarrow \Lambda^0 + \pi^+$$

which is an absorption process.

The K-mesons that exist in the experimental world(the laboratory) are linear superpositions of K^0 and $\bar{K}^0$ states in the same way that RCP and LCP photon were superpositions of $|x\rangle$ and $|y\rangle$ polarization states.

So the world of K-mesons can be represented by a 2-dimensional vector space.

One basis for the vector space is the orthonormal set $\{|K^0\rangle, |\bar{K}^0\rangle\}$ where

$$\langle K^0 | K^0 \rangle = 1 = \langle \bar{K}^0 | \bar{K}^0 \rangle \text{ and } \langle K^0 | \bar{K}^0 \rangle = 0$$

Two linear operators are important for the study of K-mesons.

First, we represent the strangeness operator.

We already stated that the states $|K^0\rangle$ and $|\bar{K}^0\rangle$ have definite values of strangeness, which means that they are eigenvectors of the strangeness operator $\hat{S}$ with eigenvalues ± 1 (by convention).

Using our formalism, this means that

$$\hat{S}|K^0\rangle = |K^0\rangle \text{ and } \hat{S}|\bar{K}^0\rangle = -|\bar{K}^0\rangle$$

$$\hat{S} = \begin{pmatrix} \langle K^0|\hat{S}|K^0\rangle & \langle K^0|\hat{S}|\bar{K}^0\rangle \\ \langle \bar{K}^0|\hat{S}|K^0\rangle & \langle \bar{K}^0|\hat{S}|\bar{K}^0\rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$
$$= |K^0\rangle\langle K^0| - |\bar{K}^0\rangle\langle \bar{K}^0|$$

in the $\{|K^0\rangle, |\bar{K}^0\rangle\}$ basis.

The second linear operator that is important in the K-meson system is charge conjugation $\hat{C}$.

This operator changes particles into antiparticles and vice versa.

In the K-meson system using the $\{|K^0\rangle, |\bar{K}^0\rangle\}$ basis we define $\hat{C}$ by the particle-antiparticle changing relations

$$\hat{C}|K^0\rangle = |\bar{K}^0\rangle \text{ and } \hat{C}|\bar{K}^0\rangle = |K^0\rangle$$

$$\hat{C} = \begin{pmatrix} \langle K^0|\hat{C}|K^0\rangle & \langle K^0|\hat{C}|\bar{K}^0\rangle \\ \langle \bar{K}^0|\hat{C}|K^0\rangle & \langle \bar{K}^0|\hat{C}|\bar{K}^0\rangle \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

We can find the eigenvectors and eigenvalues of the $\hat{C}$ operator as follows

$$\hat{C}|\psi\rangle = \lambda|\psi\rangle$$
$$\hat{C}^2|\psi\rangle = \lambda\hat{C}|\psi\rangle = \lambda^2|\psi\rangle = \hat{I}|\psi\rangle = |\psi\rangle$$

where we have used $\hat{C}^2 = \hat{I}$.

This result says that $\lambda^2 = 1$ or the eigenvalues of $\hat{C}$ are ± 1.

If we use the $\{|K^0\rangle, |\bar{K}^0\rangle\}$ basis and assume that

$$|\psi\rangle = a\left|K^0\right\rangle + b\left|\bar{K}^0\right\rangle \text{ where } |a|^2 + |b|^2 = 1$$

we find for $\lambda = +1$

$$\hat{C}|\psi\rangle = a\hat{C}\left|K^0\right\rangle + b\hat{C}\left|\bar{K}^0\right\rangle = a\left|\bar{K}^0\right\rangle + b\left|K^0\right\rangle$$
$$= |\psi\rangle = a\left|K^0\right\rangle + b\left|\bar{K}^0\right\rangle$$

or $a = b = 1/\sqrt{2}$.

If we define the $+1$ eigenvector as $|K_S\rangle$, we then have

$$|K_S\rangle = \frac{1}{\sqrt{2}}\left(\left|K^0\right\rangle + \left|\bar{K}^0\right\rangle\right)$$

Similarly, if we define the -1 eigenvector as $|K_L\rangle$, we then have

$$|K_L\rangle = \frac{1}{\sqrt{2}}\left(\left|K^0\right\rangle - \left|\bar{K}^0\right\rangle\right)$$

where

$$\hat{C}\left|K^S\right\rangle = \left|K^S\right\rangle \text{ and } \hat{C}\left|K^L\right\rangle = -\left|K^L\right\rangle$$

Since the commutator $[\hat{S}, \hat{C}] \neq 0$, these two operators do not have a common set of eigenvectors.

This means that both operators cannot have definite values in the same state.

In fact, the concept of charge conjugation is meaningless for K-mesons in the $\{|K^0\rangle, |\bar{K}^0\rangle\}$ states and the concept of strangeness is meaningless for K-mesons in the $\{|K^S\rangle, |K^L\rangle\}$ states.

The $\{|K^S\rangle, |K^L\rangle\}$ states form a second orthonormal basis for the vector space (like the RCP and LCP polarization states).

The standard approach we will follow when studying physical systems using quantum mechanics will be to

1. define the Hamiltonian for the system
2. find its eigenvalues and eigenvectors
3. investigate the time development operator generated by the Hamiltonian and
4. calculate transition probabilities connected to experiments

Along the way we will define the properties of other operators appropriate to the system under investigation (like $\hat{S}$ and $\hat{C}$ above).

It is the job of the theoretical physicist to derive or guess an appropriate Hamiltonian.

Since we are in 2-dimensional vector space, all operators are represented by 2×2 matrices.

In the case of the K-meson system, we will assume the most general form constructed from all of the relevant operators.

Therefore, we assume

$$\hat{H} = M\hat{I} + A\hat{C} + B\hat{S} = \begin{pmatrix} M+B & A \\ A & M-B \end{pmatrix}$$

and investigate the consequences of this assumption.

We will assume that the matrix has been written down in the $\{|K^0\rangle, |\bar{K}^0\rangle\}$ basis.

Step 1

Investigate the commutators:

$$[\hat{H},\hat{H}] = 0 \quad , \quad [\hat{H},\hat{S}] \neq 0 \quad , \quad [\hat{H},\hat{C}] \neq 0 \quad , \quad [\hat{S},\hat{C}] \neq 0$$

Since the Hamiltonian always commutes with itself and it is not explicitly dependent on time, the physical observable connected to the Hamiltonian, namely the energy, is conserved.

Since they do not commute with the assumed form of $\hat{H}$, neither $\hat{S}$ nor $\hat{C}$ is conserved in this model.

When a physical observable is conserved, we say its value corresponds to a good quantum number that can be used to characterize(label) the ket vector representing the physical system.

Step 2

Investigate special cases (limits of the most general solution):

Case of A = 0:

$$\hat{H} = M\hat{I} + B\hat{S} = \begin{pmatrix} M+B & 0 \\ 0 & M-B \end{pmatrix}$$

Now

$$[\hat{H},\hat{S}] = 0$$

which means that $\hat{H}$ and $\hat{S}$ share a common set of eigenvectors.

We already know the eigenvectors(non-degenerate) for $\hat{S}$ and so the eigenvector/eigenvalue problem for $\hat{H}$ is already solved(*clearly this is a very powerful rule*).

We have

$$\hat{H}\left|K^0\right\rangle = (M+B)\left|K^0\right\rangle \text{ and } \hat{H}\left|\bar{K}^0\right\rangle = (M+B)\left|\bar{K}^0\right\rangle$$

We could have surmised this from the diagonal form of the matrix representation, since the only way the matrix could be diagonal is for the basis states of the representation to be the eigenvectors.

$\hat{C}$ is not conserved in this case, since $[\hat{H}, \hat{C}] \neq 0$.

The energy eigenstates, $\{|K^0\rangle, |\bar{K}^0\rangle\}$ in this case, are a basis for the vector space.

This is always true for a Hermitian operator.

This means that we can write any arbitrary vector as a linear combination of these vectors

$$|\psi\rangle = a\left|K^0\right\rangle + b\left|\bar{K}^0\right\rangle = a\left|E = M+B\right\rangle + b\left|E = M-B\right\rangle$$

Now, as we derived earlier, energy eigenstates have a simple time dependence.

We have

$$\hat{H}\left|E\right\rangle = E\left|E\right\rangle$$

$$\hat{U}(t)\left|E\right\rangle = e^{-i\hat{H}t/\hbar}\left|E\right\rangle = e^{-iEt/\hbar}\left|E\right\rangle$$

Therefore, in this case, the time dependence of the arbitrary state vector is given by

$$|\psi(t)\rangle = ae^{-i(M+B)t/\hbar}\left|K^0\right\rangle + be^{-i(M-B)t/\hbar}\left|\bar{K}^0\right\rangle$$

This will be a general approach we will use, i.e., expand an arbitrary state in energy eigenstates and use the simple time dependence of the energy eigenstates to determine the more complex time dependence of the arbitrary state.

Of course, we have to be able to solve the eigenvector/eigenvalue problem for the Hamiltonian(the energy operator) of the system under investigation.

Case of B = 0:

$$\hat{H} = M\hat{I} + A\hat{C} = \begin{pmatrix} M & A \\ A & M \end{pmatrix}$$

Now

$$[\hat{H}, \hat{C}] = 0$$

which means that $\hat{H}$ and $\hat{C}$ share a common set of eigenvectors.

We already know the eigenvectors(non-degenerate) for C and so the eigenvector/eigenvalue problem for $\hat{H}$ is again already solved.

We have

$$\hat{H}\,|K_S\rangle = (M + A)\,|K_S\rangle \text{ and } \hat{H}\,|K_L\rangle = (M - A)\,|K_L\rangle$$

$\hat{S}$ is not conserved in this case, since [$\hat{H}$, $\hat{S}$] $\neq 0$.

The energy eigenstates, $\{|K_S\rangle, |K_L\rangle\}$ in this case, are a basis for the vector space.

In this basis

$$\hat{H} = \begin{pmatrix} M + A & 0 \\ 0 & M - A \end{pmatrix}$$

as expected.

We could also solve this problem by finding the characteristic equation for the Hamiltonian matrix, i.e., since we have

$$\hat{H}\,|\psi\rangle = \lambda\,|\psi\rangle \rightarrow (\hat{H} - \lambda\hat{I})\,|\psi\rangle = 0$$

the characteristic equation is

$$det(\hat{H} - \lambda\hat{I}) = det\begin{pmatrix} M-\lambda & A \\ A & M=\lambda \end{pmatrix}$$

$$= (M-\lambda)^2 - A^2 = 0 \rightarrow \lambda = M \pm A$$

Since we have another basis, we can write any arbitrary vector as a linear combination of these vectors

$$|\psi\rangle = a\,|K_S\rangle + b\,|K_L\rangle = a\,|E = M + A\rangle + b\,|E = M - A\rangle$$

Therefore, in this case, we have the time dependence

$$|\psi(t)\rangle = ae^{-i(M+A)t/\hbar}\,|K_S\rangle + be^{-i(M-A)t/\hbar}\,|K_L\rangle$$

Step 3

Solve the general Hamiltonian problem (if possible; otherwise we must use approximation methods, which we will discuss later).

We have

$$\hat{H} = M\hat{I} + A\hat{C} + B\hat{S} = \begin{pmatrix} M+B & A \\ A & M-B \end{pmatrix}$$

We assume that the eigenvectors satisfy $\hat{H}|\phi\rangle = E\,|\phi\rangle$ where

$$|\phi\rangle = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

This gives

$$\begin{pmatrix} M+B & A \\ A & M-B \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = E\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

or

$$\begin{pmatrix} M+B-E & A \\ A & M-B-E \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = 0$$

This is a set of two homogeneous equations in two unknowns.

It has a nontrivial solution only if the determinant of the coefficients is zero

$$\begin{vmatrix} M+B-A & A \\ A & M-B-E \end{vmatrix} = (M+B-E)(M-B-E) - A^2 = 0$$

This has solution

$$E_\pm = M \pm \sqrt{A^2 + B^2} \quad \text{(the energy eigenvalues)}$$

We solve for the eigenstates by substituting the eigenvalues into the eigenvalue/eigenvector equation

$$\begin{pmatrix} M+B & A \\ A & M-B \end{pmatrix}\begin{pmatrix} \phi_{1\pm} \\ \phi_{2\pm} \end{pmatrix} = E_\pm\begin{pmatrix} \phi_{1\pm} \\ \phi_{2\pm} \end{pmatrix}$$

After some algebra we get

$$\frac{\phi_{1\pm}}{\phi_{2\pm}} = \frac{-A}{\sqrt{B \pm A^2 + B^2}} = \frac{B \pm \sqrt{A^2 + B^2}}{A}$$

We check the validity of this solution by comparing it the limiting cases; that is why we looked at the special cases earlier.

For $B = 0$, we have

$$\frac{\phi_{1\pm}}{\phi_{2\pm}} = \frac{-A}{\mp A} = \pm 1 \rightarrow \phi_{1+} = \phi_{2+} \text{ and } \phi_{1-} = -\phi_{2-}$$

which says that

$$|\phi_+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} = |K_S\rangle \text{ and } |\phi_+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} = |K_L\rangle$$

which agrees with the earlier results for this case.

In the other limiting case, $A = 0$, we have

$$\frac{\phi_{1+}}{\phi_{2+}} = \infty \rightarrow \phi_{1+} = 1 \text{ and } \phi_{2+} = 0$$

$$\frac{\phi_{1-}}{\phi_{2-}} = 0 \rightarrow \phi_{1-} = 0 \text{ and } \phi_{2-} = 1$$

which says that

$$|\phi_+\rangle = \begin{pmatrix}1\\0\end{pmatrix} = |K^0\rangle \text{ and } |\phi_+\rangle = \begin{pmatrix}0\\1\end{pmatrix} = |\bar{K}^0\rangle$$

which again agrees with the earlier results for this case.

If we normalize the general solution

$$\frac{\phi_{1\pm}}{\phi_{2\pm}} = \frac{-A}{\sqrt{B \pm A^2 + B^2}} = \frac{B \pm \sqrt{A^2 + B^2}}{A}$$

using

$$|\phi_{1\pm}|^2 + |\phi_{2\pm}|^2 = 1$$

we obtain

$$\phi_{1\pm} = \frac{A}{\sqrt{A^2 + \left(B \mp \sqrt{A^2+B^2}\right)^2}}$$

$$\phi_{2\pm} = \frac{B \mp \sqrt{A^2+B^2}}{\sqrt{A^2 + \left(B \mp \sqrt{A^2+B^2}\right)^2}}$$

and

$$|\phi_\pm\rangle = \frac{1}{\sqrt{A^2 + \left(B \mp \sqrt{A^2+B^2}\right)^2}} \begin{pmatrix} A \\ -B \pm \sqrt{A^2+B^2} \end{pmatrix}$$

$$= \frac{1}{\sqrt{A^2 + \left(B \mp \sqrt{A^2+B^2}\right)^2}} \left[A|K^0\rangle + \left(-B \pm \sqrt{A^2+B^2}\right)|\bar{K}^0\rangle \right]$$

$$= \frac{1}{2\sqrt{A^2 + \left(B \mp \sqrt{A^2+B^2}\right)^2}} \times$$

$$\left[\left(A - B \pm \sqrt{A^2+B^2}\right)|K_S\rangle + \left(A + B \pm \sqrt{A^2+B^2}\right)|K_L\rangle \right]$$

Step 4

Look at a realistic physical system that we can relate to experiment.

In the real world of K-mesons, the Hamiltonian is such that B ≪ A. In this case the states $\{|K_S\rangle, |K_L\rangle\}$ are *almost* energy eigenstates or charge conjugation is *almost* conserved.

We expect that instead of being able to write

$$|\phi_+\rangle = |K_S\rangle \text{ and } \phi_- = |K_L\rangle$$

which would be true if $B = 0$, we should be able to write

$$|\phi_+\rangle = \cos\frac{\theta}{2}|K_S\rangle + \sin\frac{\theta}{2}|K_L\rangle$$
$$|\phi_-\rangle = -\sin\frac{\theta}{2}|K_S\rangle + \cos\frac{\theta}{2}|K_L\rangle$$

where for $\theta \ll 1$ we clearly approximate the $B = 0$ result.

Let us see how this works.

For $B \ll A$, we choos

$$\frac{\theta}{2} = \frac{B}{2A} \ll 1$$

and get

$$\frac{\theta}{2} \approx \tan\frac{\theta}{2} = \frac{B}{2A} = \frac{\sin\frac{\theta}{2}}{\cos\frac{\theta}{2}}$$

To lowest order we can then say

$$\sin\frac{\theta}{2} = \frac{B}{2A} = \delta \ll 1 \text{ and } \cos\frac{\theta}{2} = 1$$

to get

$$|\phi_+\rangle = |K_S\rangle + \delta|K_L\rangle$$
$$|\phi_-\rangle = |K_L\rangle - \delta|K_S\rangle$$

This says that if $|\psi_{\text{in}}\rangle = |\phi_-\rangle$, then the number

$$\frac{|\langle K_S|\phi_-\rangle|^2}{|\langle K_L|\phi_-\rangle|^2} = \frac{\text{probability of observing a } K_S}{\text{probability of observing a } K_L}$$
$$= \frac{(\text{Number of } K_S)/(\text{Number of } K_S \text{ and } K_L)}{(\text{Number of } K_L)/(\text{Number of } K_S \text{ and } K_L)}$$
$$= \frac{(\text{Number of } K_S)}{(\text{Number of } K_L)}$$

gives the experimental ratio of the number of times we will measure a final state of $|K_S\rangle$ to the number of times we will measure the final state of $|K_L\rangle$.

The signature for seeing a final state of $|K_L\rangle$ is to see a decay to 3 π-mesons and that of a final state of $|K_S\rangle$ is to see a decay to 2 π-mesons.

The number is

$$\frac{|\langle K_S|\phi_-\rangle|^2}{|\langle K_L|\phi_-\rangle|^2} = |\delta|^2$$

Now experiment gives the result $|\delta| = 2 \times 10^{-3}$.

This number is a measure of how large of an effect strangeness non-conservation has on this system.

If $B = 0$, then charge conjugation is conserved.

If $B \neq 0$, then charge conjugation is not absolutely conserved.

So δ is a measure of the lack of charge conjugation conservation in the K-meson system.

If we identify the energy eigenvalues as the particle *rest energies*

$$M + A = m_S c^2 \quad , \quad M - A = m_L c^2$$

we then have

$$A = \frac{m_S c^2 - m_L c^2}{2} = 10^{-5}\, eV$$

and

$$2A\delta = 10^{-17} m_K c^2$$

or

$$\frac{B}{m_K c^2} = 10^{-17}$$

Thus, this is a one part in 10^{-17} effect!

It is one of the best measurements ever made.

It won the Nobel prize for the experimenters in 1963.

It is now understood in detail by the standard model of elementary particles.

Now let us look at **Quantum Interference Effects** in this K-meson system.

Suppose that $B = 0$.

Then the energy eigenstates are $\{|K_S\rangle, |K_L\rangle\}$ with eigenvalues $M \pm A$.

Now let

$$|\psi_{in}\rangle = |K^0\rangle = \frac{1}{\sqrt{2}}(|K_S\rangle + |K_L\rangle)$$

This is not an energy eigenstate so it will evolve in time.

Its time evolution is given by

$$\begin{aligned}|\psi(t)\rangle &= e^{-i\hat{H}t/\hbar}\,|\psi_{in}\rangle = e^{-i\hat{H}t/\hbar}\,\left|K^0\right\rangle \\ &= \frac{1}{\sqrt{2}}(e^{-i\hat{H}t/\hbar}\,|K_S\rangle + e^{-i\hat{H}t/\hbar}\,|K_L\rangle) \\ &= \frac{1}{\sqrt{2}}(e^{-i(M+A)t/\hbar}\,|K_S\rangle + e^{-i(M-A)t/\hbar}\,|K_L\rangle)\end{aligned}$$

The probability amplitude that the initial meson *changes(oscillates)* into the orthogonal state $|\,\bar{K}^0\rangle$ at time *t* is given by

$$\begin{aligned}\left\langle \bar{K}^0 \middle| \psi(t) \right\rangle &= \frac{1}{\sqrt{2}}(e^{-i(M+A)t/\hbar}\left\langle \bar{K}^0 \middle| K_S \right\rangle + e^{-i(M-A)t/\hbar}\left\langle \bar{K}^0 \middle| K_L \right\rangle) \\ &= \frac{1}{2}(e^{-i(M+A)t/\hbar} - e^{-i(M-A)t/\hbar})\end{aligned}$$

Finally, the probability that the incoming K^0-meson will behave like a $\bar{K}^0$-meson at time *t* (that it has oscillated into a $\bar{K}^0$) is given by

$$P_{\bar{K}^0}(t) = |\left\langle \bar{K}^0 \middle| \psi(t) \right\rangle|^2 = \frac{1}{2}[1 - \cos\Omega t]$$

where

$$\Omega = \frac{1}{\hbar}((M + A) - (M - A)) = \frac{2A}{\hbar} = \frac{m_S c^2 - m_L c^2}{\hbar}$$

What is the physics here?

If $m_S - m_L = 0$, then $\Omega = 0$ and $P_{\bar{K}^0}(t) = 0$ or the two mesons do not change into one another(called an *oscillation*) as time passes.

However, if If $m_S - m_L \neq 0$, then $P_{\bar{K}^0}(t) \neq 0$ and the two mesons oscillate back and forth, sometimes being a K^0 and sometimes being a $\bar{K}^0$.

This has been observed in the laboratory and is, in fact, the way that the extremely small mass difference is actually measured.

This is also the same mechanism that operates for oscillations between the different flavors of neutrinos.

Experiment now indicate that neutrino oscillation is taking place and the explanation is that the neutrinos have different masses, some non-zero!

Stern-Gerlach Experiments and Measurement

From early in our exploration of quantum mechanics, measurement played a central role.

A basic axiom of standard quantum theory is the the **von Neumann projection postulate**:

> If an observable $\hat{A}$ is measured, the result is one of its eigenvalues, a. After the measurement, the system is *projected* into the eigenvector $|a\rangle$. If $|\psi\rangle$ is the state before the measurement, the probability of this occurence is $|\langle a|\psi\rangle|^2$.

Let us now explore the physical consequences of these measurements.

The most fundamental example is the measurement of the angular momentum component of a spin-1/2 particle which takes on only two possible values.

We will discuss the full theory of spin in later.

The measurement was first carried out by Stern and Gerlach in 1922 to test Bohr's ideas about "space quantization".

This was before Uhlenbeck and Goudsmidt's invention of spin angular momentum, so Stern and Gerlach's results were not completely understood.

Nonetheless, the results were startling and one of the first real pictures of the strange quantum world.

Consider the force on a magnetic moment which moves through a spatially inhomogeneous magnetic field.

The potential energy is

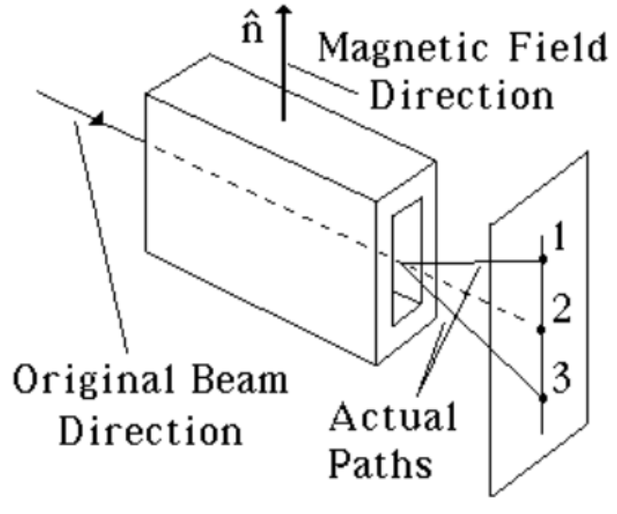
$$V = \vec{\mu} \cdot \vec{B}(\vec{r})$$

and thus the force is

$$\vec{F} = -\nabla V = \nabla(\vec{\mu} \cdot \vec{B}(\vec{r})) = (\vec{\mu} \cdot \nabla)\vec{B}(\vec{r})$$

Stern and Gerlach set up a spatially inhomogeneous field with one very large component(call it the z-component).

A schematic diagram of a Stern-Gerlach apparatus is shown in Figure below.



Stern-Gerlach Apparatus

We then have

$$\hat{n} = \hat{z} \quad , \quad \vec{B} \propto B(z)\hat{z} + \text{ small x,y components}$$

$$\Rightarrow (\vec{\mu} \cdot \nabla)\vec{B}(\vec{r}) \propto \mu_z \frac{\partial B}{\partial z}\hat{z}$$

Now quantum mechanically, $\vec{\mu} = \gamma_J \vec{J}$ which gives

$$\vec{F} = \left(\gamma_J \frac{\partial B}{\partial z} J_z \hat{Z}\right)$$

Stern and Gerlach sent Na (sodium) atoms from an oven, collimated into a beam, into the region of ingomogeneous field.

Na is an alkali atom with one valence electron with orbital angular momentum $\ell = 0$ (an s-state).

Thus, all of the angular momentum of the atom is due to the spin angular momentum of the single valence electron (spin 1/2).

Thus

$$\vec{\mu} = -\frac{2\mu_B}{\hbar}\vec{S} = -\mu_B \vec{\sigma} \Rightarrow \vec{F} = -\mu_B \frac{\partial B}{\partial z}\sigma_z \hat{z}$$

Clearly, the spin-up atoms $|\uparrow_x\rangle$ will experience a different interaction than spin- down atoms $|\downarrow_x\rangle$ (due to two different eigenvalues of σ_z), thereby splitting the beam into two spatially separated beams as shown in the Figure above.

We note that the spins which emerge from the oven have a random orientation.

In general, a Stern-Gerlach apparatus takes an incoming beam of particles with angular momentum and splits the beam into a number of spots on a screen.

The number of spots is equal to 2J + 1, where J = angular momentum value of the incoming beam.

We shall see in later chapters that 2J +1 = the number of values allowed quantum mechanically for the measurable quantity $\mathbf{J} \cdot \hat{\mathbf{n}}$ where $\hat{\mathbf{n}}$ is unit vector in the direction of the magnetic field inside the apparatus.

In the diagram, spot #2 (undeflected beam) is where the beam would have hit the screen if no magnetic field were present in the apparatus.

Spots #1 and #3 are an example the 2J + 1 = 2 spots we would observe in an experiment when J = 1/2 (the original Stern-Gerlach experiment).

The important features of the apparatus for our theoretical discussion are:

1. The breakup into a finite number of discrete beams (we will assume we are working with $J = 1/2$ particles and thus have 2 beams exiting the apparatus)
2. The beams are *separated in 3-dimensional space* and each contains 1/2 of the original particles entering the device
3. The possible values of $\mathbf{B} \cdot \hat{\mathbf{n}}$ for $J = 1/2$ are $\pm\hbar/2$ for any $\hat{\mathbf{n}}$.
4. One exiting beam contains *only* particles with $\mathbf{J} \cdot \hat{\mathbf{n}} = +\hbar/2$ and the other beam contains *only* particles with $\mathbf{J} \cdot \hat{\mathbf{n}} = -\hbar/2$

   We will represent the beam state vectors by the ket vectors

$$|+\hat{\mathbf{n}}\rangle = \text{ beam with } \mathbf{J} \cdot \hat{\mathbf{n}} = +\hbar/2$$
$$|-\hat{\mathbf{n}}\rangle = \text{ beam with } \mathbf{J} \cdot \hat{\mathbf{n}} = -\hbar/2$$

   and the Stern-Gerlach apparatus with a field in the $\hat{\mathbf{n}}$-direction by SG$\hat{\mathbf{n}}$.
5. The above results occur for all beams no matter what the direction of the unit vector $\hat{\mathbf{n}}$

The S-G experiment has all the essential ingredients of a quantum measurement.

The quantum *degree of freedom*, here spin-1/2, is correlated with the final beam direction of the atoms.

Once the two spots on the screen are *resolvable*, we can measure the spin state of the atom by determining which spatially separated beam the atom is in.

We *now report the results of a series of actual experiments*.

Experiment #1

We send N particles into an SG$\hat{\mathbf{z}}$ device and select out the beam where the particles are in the state $|+\hat{\mathbf{z}}\rangle$ (we block the other beam).

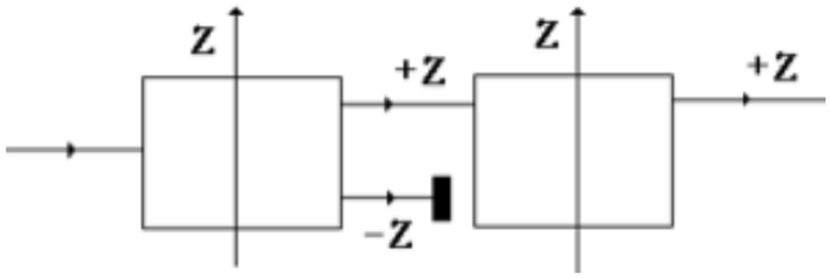
It contains $N/2$ particles.

We then send this second beam into another SG$\hat{\mathbf{z}}$ device.

We find that all $N/2$ exit in the state $|+\hat{\mathbf{z}}\rangle$.

There is only one exit beam.

Symbolically, this looks like Figure below:



Experiment #1

This says that when we make a measurement, say $\mathbf{J}\cdot\hat{\mathbf{z}}$ and then immediately make another measurement of the same quantity, we get the same result as first measurement with probability = 1.

Since the only way to a measurement result with certainty is to be in an eigenstate of the observable, the measurement seems to have caused the system to change from a superposition of eigenstates into two beams(separated in physical space) each with a definite state of the observable $\mathbf{J} \cdot \hat{\mathbf{z}}$ (one of its eigenstates).

This experiment is called **state preparation**.

We prepare a state by measuring an observable and finding one of the eigenvalues.

Afterwards, by the *von Neumann projection postulate*, the state is the corresponding eigenvector.

In the above experiment, by redirecting the *spin-up* beam into the second SG apparatus we are guaranteed to find the eigenstate prepared by the first SG apparatus (with the block).

Thus a screen would have only one spot after the second apparatus.

Experiment 2

We send N particles into an SG $\hat{\mathbf{z}}$ device and select out the beam where the particles are in the state $|+\ \hat{\mathbf{z}}\ \rangle$.

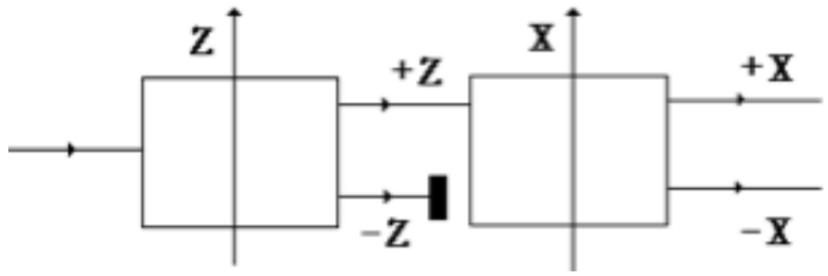
It contains $N/2$ particles.

We then send the selected beam into an SG $\hat{\mathbf{x}}$ device.

We find that $N/4$ exit in the state $|+\hat{\mathbf{x}}\ \rangle$ and $N/4$ exit in the state $|-\ \hat{\mathbf{x}}\ \rangle$.

At the end, there are two exit beams.

Symbolically, this looks like the Figure:



Experiment #2

The same thing happens if we stop the $|+\hat{\mathbf{z}}\rangle$ beam and let the $|-\hat{\mathbf{z}}\rangle$ beam into the SG $\hat{\mathbf{x}}$ device.

So an SG $\hat{\mathbf{x}}$ device takes a beam with a definite value of $\mathbf{J}\cdot\hat{\mathbf{z}}$ and randomizes it, i.e., we once again have two exiting beams with equal numbers of particles.

This is saying that for a system in an eigenstate of one observable, the measurement of an incompatible observable (observables do not commute) randomizes the value of the original observable.

In this case $[\,\hat{J}_z, \hat{J}_x\,] \neq 0$].

As we will show later, the $|+\hat{\mathbf{z}}\rangle$ state is not an eigenstate of $\hat{J}_x$.

In fact,

$$|+\hat{\mathbf{z}}\rangle = \frac{1}{\sqrt{2}}(|+\hat{\mathbf{x}}\rangle + |-\hat{\mathbf{x}}\rangle)$$

i.e., a 50-50 superposition of spin-up and spin-down in the x-direction.

Thus, we see two beams after the SG $\hat{\mathbf{x}}$ apparatus.

So even though a pure state is entering the SG $\hat{\mathbf{x}}$ apparatus, we do not get a definite value because $[\,\hat{J}_z, \hat{J}_x\,] \neq 0$].

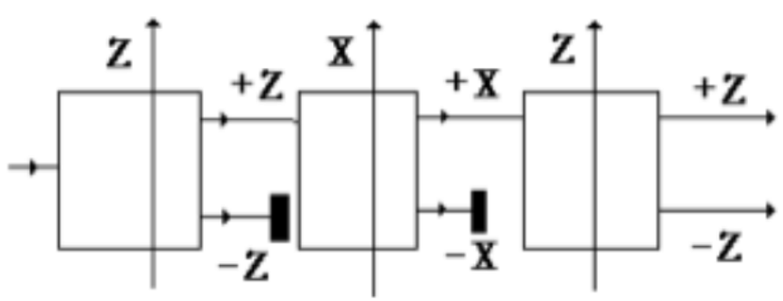
Experiment 3

We now add a third SG device to Experiment 2.

It is an $|+\hat{\mathbf{z}}\rangle$ device.

We also block the $|+\hat{\mathbf{x}}\rangle$ exiting

beam as shown symbolically in the Figure.



Experiment #3

We found above that $N/4$ exited in the state $|+\,\hat{\mathbf{x}}\,\rangle$ from the SG $\hat{\mathbf{x}}$ device.

After the third device we find that $N/8$ exit in the state $|+\,\hat{\mathbf{z}}\,\rangle$ and $N/8$ exit in the state $|-\,\hat{\mathbf{z}}\,\rangle$.

What has happened?

It seems that making a measurement of $\mathbf{J}\cdot\hat{\mathbf{x}}$ on a beam with definite $\mathbf{J}\cdot\hat{\mathbf{z}}$ modifies the system rather dramatically.

We did two successive measurements on these particles.

Since we isolated the + beam in each case we might be led to think that the beam entering the last SG $\hat{\mathbf{z}}$ device (because of our selections) has

$$\mathbf{J}\cdot\hat{\mathbf{z}} = +\frac{\hbar}{2} \text{ AND } \mathbf{J}\cdot\hat{\mathbf{x}} = +\frac{\hbar}{2}$$

But the experiment says this cannot be so, since 50% of the particles exiting the last device have

$$\mathbf{J}\cdot\hat{\mathbf{z}} = -\frac{\hbar}{2}$$

We are forced to say that the SG $\hat{\mathbf{z}}$ device takes a definite value of $\mathbf{J}\cdot\hat{\mathbf{x}}$ and randomizes it so that we end up with two exiting beams with equal numbers of particles.

Why?

Again, since $[\hat{J}_x, \hat{J}_z] \neq 0$ the two observables cannot share a common set of eigenvectors or they are *incompatible*.

Since an observable can only have a definite value when the state is one of its eigenvectors, it is not possible for both $\mathbf{J}\cdot\hat{\mathbf{z}}$ and $\mathbf{J}\cdot\hat{\mathbf{x}}$ to simultaneously have definite values.

Our two successive measurements **DO NOT** produce definite values for both observables.

Each measurement only produces a definite value for the observable it is measuring and randomizes the incompatible observable (actually randomizes all other incompatible observables)!

All measurement results depend on the *context* of the measurement.

Another way to think about this is to say the following.

An SG $\hat{\mathbf{z}}$ device is a measurement of the angular momentum in the $\hat{\mathbf{z}}$-direction.

Any such measurement of an observable randomizes the next measurement of any other incompatible observable.

In this cascaded measurement, the probability of finding $|\downarrow\rangle$ is the product of conditional probabilities for uncorrelated events

$$P(\downarrow_z \mid \uparrow_x, \uparrow_z) = P(\downarrow_z \mid \uparrow_x) P(\downarrow_z \mid \uparrow_z) = |\langle \downarrow_z | \uparrow_x \rangle|^2 |\langle \uparrow_x | \uparrow_z \rangle|^2 = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

as is observed.

Now without the intermediate apparatus, *classical theory* would argue as follows.

If we don't measure J_x we must sum the probabilities of all possible alternatives or

$$P(\downarrow_z \mid \uparrow_z) = P(\downarrow_z \mid \uparrow_x) P(\uparrow_x \mid \uparrow_z) + P(\downarrow_z \mid \downarrow_x) P(\downarrow_x \mid \uparrow_z) = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}$$

which differs from the experimental result(= 0).

As we have discussed, quantum mechanically, if the different possibilities are indistinguishable (i.e., there is no information available to distinguish the different alternatives) we must add the probability amplitudes (before squaring) so that these different alternatives can interfere.

$$
\begin{aligned}
P(\downarrow_z \mid \uparrow_z) &= |\langle \downarrow_z | \uparrow_z \rangle|^2 = |\langle \downarrow_z | \hat{I} | \uparrow_z \rangle|^2 \\
&= |\langle \downarrow_z | \left(|\uparrow_x\rangle \langle \uparrow_x| + |\downarrow_x\rangle \langle \downarrow_x| \right) | \uparrow_z \rangle|^2 \\
&= \underbrace{P(\downarrow_z \mid \uparrow_x) P(\uparrow_x \mid \uparrow_z) + P(\downarrow_z \mid \downarrow_x) P(\downarrow_x \mid \uparrow_z)}_{classical\ terms} + \text{interference terms} \\
&= \frac{1}{2} - \frac{1}{2} = 0
\end{aligned}
$$

So, the quantum amplitude interference calculation gives the correct result.

What constitutes a measurement?

If we send atoms with spin-up in the z-direction into an x-oriented SG apparatus, there is a $50-50$ probability of emerging as spin-up or spin-down in the x-direction.

But if we do not detect which port the atom exits from, did we measure the spin in the x-direction?

In a sense, no.

A measurement is something that removes coherence, i.e., the ability for quantum processes to interfere.

In principle, we can recombine these beams as if no measurement had occurred, i.e., we can still have interference - we do not remove coherence unless we look(something permanent happens!).

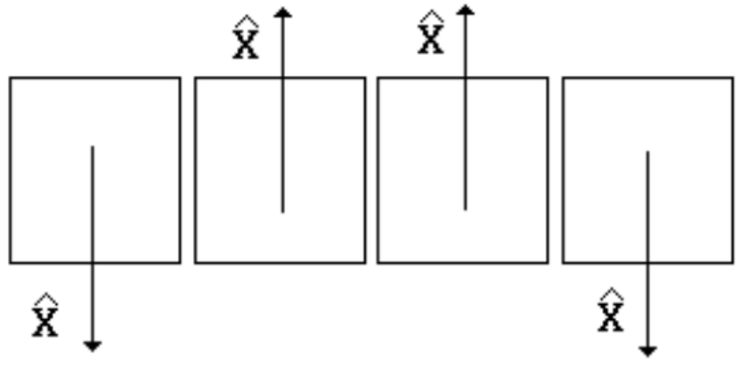
Again the context of the experiment is the significant feature!

Experiment 4

Now, let us construct a new device to illustrate this point.

It is called a modified SG$\hat{\mathbf{x}}$ device.

It looks like the Figure.



Experiment #4

Any beam of particles entering this modified device would experience deflections while traveling though it.

However, the device lengths and field strengths have been cleverly chosen so that the *net effect for any beam is no change!*

The internal beams are all recombined so that the state of the beam upon exiting the entire device is identical to the state of the beam before entering the device.

This device might be called a *total-of-nothing* device.

It turns out, however, that we can use this device to make a measurement and select a particular spin state.

We can calculate the paths that would be followed in the device by the $|+\hat{\mathbf{x}}\rangle$ and $|-\hat{\mathbf{x}}\rangle$ (these are the relevant beams to consider because these are SG $\hat{\mathbf{x}}$ devices).

Using this information, we can block the path that a particle in the state $|-\hat{\mathbf{x}}\rangle$ would follow.

Then, all the particles exiting the modified SG $\hat{\mathbf{x}}$ device would be in the state $|+\hat{\mathbf{x}}\rangle$.

Experiment 5

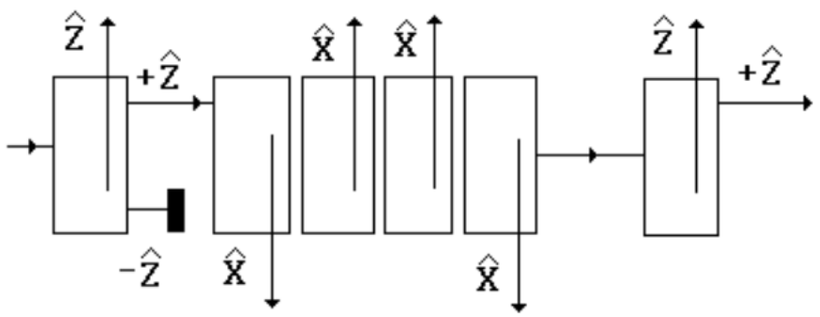
We can confirm this fact by inserting a modified SG$\hat{\mathbf{x}}$ device into Experiment 3 where we replace the SG$\hat{\mathbf{x}}$ device by the modified SG$\hat{\mathbf{x}}$ device.

If we block the $|-\hat{\mathbf{x}}\rangle$ beam, the state at the end of the modified SG$\hat{\mathbf{x}}$ device is the same as it was after the original SG$\hat{\mathbf{x}}$ device and we get exactly the same results.

In fact, we get the same result whether we block the $|-\hat{\mathbf{x}}\rangle$ beam or the $|+\hat{\mathbf{x}}\rangle$ beam, as should be the case.

Now we set up Experiment 5.

It is shown in the Figure.



Experiment #5

In this experiment a beam enters the SG$\hat{\mathbf{z}}$ device and we block the exiting $|-\hat{\mathbf{z}}\rangle$ beam.

The $|+\hat{\mathbf{z}}\rangle$ beam is sent into the modified SG$\hat{\mathbf{x}}$ device and the exit beam is sent into the final SG$\hat{\mathbf{z}}$ device.

In this case, however, we DO NOT block any of the paths in the modified SG$\hat{\mathbf{x}}$ device.

Since the beam entering the modified SG$\hat{\mathbf{x}}$ device is reconstructed before it exits (we already saw that it does this in the Experiment 4), we are NOT making a measurement of $\mathbf{J}\cdot\hat{\mathbf{x}}$ using the modified SG$\hat{\mathbf{x}}$ device as we did when we used the original SG$\hat{\mathbf{x}}$ device.

Now we send in N particles and $N/2$ are in the $|+\hat{\mathbf{z}}\rangle$ beam as before.

However, now, instead of find $N/8$ particles in the final $|+\hat{\mathbf{z}}\rangle$ beam, we find $N/2$ particles.

ALL the particles make it through unchanged, even though there are SG$\hat{\mathbf{x}}$ devices in between.

It behaves as if the modified SG$\hat{\mathbf{x}}$ device was not there at all(hence its name).

We might have assumed that 50% of the particles in the $|+\,\hat{\mathbf{z}}\,\rangle$ beam before the modified SGx^ device would emerge in the $|+\,\hat{\mathbf{x}}\,\rangle$ state and the other 50% would emerge in the $|-\,\hat{\mathbf{x}}\,\rangle$ state.

Experiment 5 says this cannot be true, since if it were true, then we would have two beams coming out of the final SG $\hat{\mathbf{z}}$ device, each with 50% of the particles.

Our results are incompatible with the statement that the particles passing through the modified SG $\hat{\mathbf{x}}$ device are either in the state $|+\,\hat{\mathbf{x}}\,\rangle$ or in the state $|-\,\hat{\mathbf{x}}\,\rangle$.

In fact, if we carry out this experiment with a very low intensity beam where only one particle at a time is passing through the apparatus, then we observe that each particle emerging from the final SG $\hat{\mathbf{z}}$ device is in the state $|+\,\hat{\mathbf{z}}\,\rangle$.

This eliminates any explanation of the result that would invoke interactions among the particles while they were in the apparatus.

For beams of particles, we have been talking in terms of percentages or fractions of the particles as experimental results.

For a single particle, however, it is not possible to predict with certainty the outcome of a measurement in advance.

We have seen this earlier in our experiments with photons and polaroids.

We are only able to use probability arguments in this case.

In Experiment 2, for instance, before a measurement (passing through the SGx^ device) of $\mathbf{J}\cdot\hat{\mathbf{x}}$ on a single particle in the $|+\,\hat{\mathbf{z}}\,\rangle$ state, all we can say is that there is a 50% probability of obtaining $|+\hat{\mathbf{x}}\,\rangle$ and a 50% probability of obtaining $|-\,\hat{\mathbf{x}}\,\rangle$.

Probabilities alone are not enough, however, to explain Experiment 5.

We came to the incorrect conclusion because we made the same mistake as in our earlier discussion of polarization.

We added the separate probabilities of the indistinguishable ways to get the total probability, whereas the correct result, as we know, is to add the amplitudes of the indistinguishable ways and then square the total amplitude to get the correct result.

We eliminated the interference effects!

When we don't actually make a measurement as in the modified SG $\hat{\mathbf{x}}$ device, we must add amplitudes and not probabilities.

We can now introduce a formalism, which should allow us to explain all experiments correctly.

We need a 2-dimensional vector space to describe these physical systems.

As a basis for this space we can use any of the sets $\{|+\hat{\mathbf{n}}\rangle, |-\hat{\mathbf{n}}\rangle\}$ corresponding to the definite values $\pm\hbar/2$ for $\mathbf{J}\cdot\hat{\mathbf{n}}$.

Each of these is an orthonormal basis where

$$\langle +\hat{\mathbf{n}}|+\hat{\mathbf{n}}\rangle = 1 = \langle -\hat{\mathbf{n}}|-\hat{\mathbf{n}}\rangle \text{ and } \langle +\hat{\mathbf{n}}|-\hat{\mathbf{n}}\rangle = 0$$

Any arbitrary state can be written as a superposition of the basis states

$$|\psi\rangle = \langle +\hat{\mathbf{n}}|\psi\rangle\,|+\hat{\mathbf{n}}\rangle + \langle -\hat{\mathbf{n}}|\psi\rangle\,|-\hat{\mathbf{n}}\rangle$$

and the operators can be written as

$$\mathbf{J}\cdot\hat{\mathbf{n}} = \frac{\hbar}{2}|+\hat{\mathbf{n}}\rangle\langle +\hat{\mathbf{n}}| - \frac{\hbar}{2}|-\hat{\mathbf{n}}\rangle\langle -\hat{\mathbf{n}}|$$

Finally, expectation values are given by

$$\langle \hat{J}_z \rangle = \langle \psi | \hat{J}_z | \psi \rangle = \frac{\hbar}{2} | \langle +\hat{\mathbf{z}} | \psi \rangle |^2 - \frac{\hbar}{2} | \langle -\hat{\mathbf{z}} | \psi \rangle |^2$$

Analysis of Experiment 3

In Experiment 3, the state before entering the first SG $\hat{\mathbf{z}}$ device is

$$|\psi_1\rangle = a |+\hat{\mathbf{z}}\rangle + b |-\hat{\mathbf{z}}\rangle \text{ where } |a|^2 + |b|^2 = 1$$

Since the SG $\hat{\mathbf{z}}$ device is a measurement of $\hat{J}_z$, after the SG $\hat{\mathbf{z}}$ device the state is $|+\hat{\mathbf{z}}\rangle$ (remember the we blocked the $|-\hat{\mathbf{z}}\rangle$ path).

It is very important to realize that we cannot answer a question about a measurement unless we express the state in the basis consisting of the eigenvectors of the operator representing the observable being measured.

We call this the *home space* of the observable.

That is why we used the $\hat{J}_z$ eigenvectors to discuss the measurement made using the SG $\hat{\mathbf{z}}$ device. Now, we are going to make a measurement of $\hat{J}_x$ in the SG $\hat{\mathbf{x}}$ device.

So we now *switch* to a basis consisting of the $\hat{J}_x$ eigenvectors.

We know we can write

$$|+\hat{\mathbf{x}}\rangle = \langle +\hat{\mathbf{z}} | +\hat{\mathbf{x}} \rangle |+\hat{\mathbf{z}}\rangle + \langle -\hat{\mathbf{z}} | +\hat{\mathbf{x}} \rangle |-\hat{\mathbf{z}}\rangle$$

One of our experiments tells us that when we send a particle in a $|+\hat{\mathbf{x}}\rangle$ state through an SG $\hat{\mathbf{z}}$ device, the probability = 1/2 that we find $|+\hat{\mathbf{z}}\rangle$ and 1/2 that we find $|-\hat{\mathbf{z}}\rangle$.

This means that

$$|\langle +\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle|^2 = |\langle -\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle|^2 = \frac{1}{2}$$

or

$$\langle +\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle = \frac{e^{1\alpha_+}}{\sqrt{2}} \text{ and } \langle -\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle = \frac{e^{1\alpha_+}}{\sqrt{2}}$$

so that

$$|+\hat{\mathbf{x}}\rangle = \frac{e^{1\alpha_+}}{\sqrt{2}}|+\hat{\mathbf{z}}\rangle + \frac{e^{1\alpha_+}}{\sqrt{2}}|-\hat{\mathbf{z}}\rangle$$

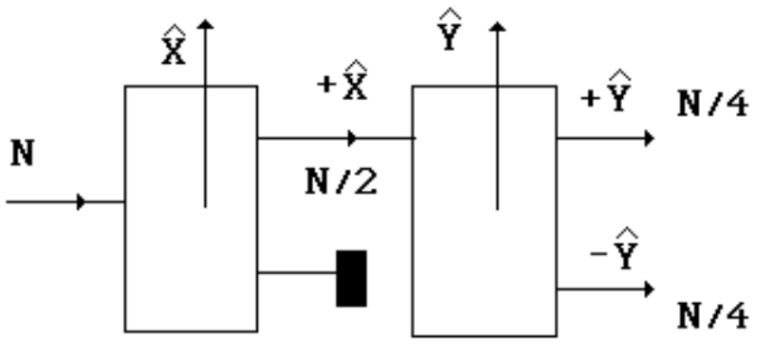
Similarly,

$$|-\hat{\mathbf{x}}\rangle = \frac{e^{1\beta_+}}{\sqrt{2}}|+\hat{\mathbf{z}}\rangle - \frac{e^{1\beta_+}}{\sqrt{2}}|-\hat{\mathbf{z}}\rangle$$

Experiment 6

In this experiment we replace the last SG $\hat{\mathbf{z}}$ device in Experiment 3 with an SG $\hat{\mathbf{y}}$ device.

The last part of the setup is shown in the Figure.



Experiment #6

The incoming beam is from the first SG $\hat{\mathbf{z}}$ device, where we blocked the $|-\hat{\mathbf{z}}\rangle$ path, and therefore, we have *N* particles in the state $|+\hat{\mathbf{z}}\rangle$.

They now go through the SG $\hat{\mathbf{x}}$ device, where we block the $|-\hat{\mathbf{x}}\rangle$ path, as shown.

The beam entering the SG $\hat{\mathbf{y}}$ device has $N/2$ particles in the $|+\hat{\mathbf{x}}\rangle$ state.

Whether we call the last direction the z-direction or the y-direction cannot affect the results of the experiment, so we get the same result for this experiment as we did for Experiment 3, except that we use the y-label instead of the z-label.

If the SG $\hat{\mathbf{x}}$ device were replaced by an SG $\hat{\mathbf{z}}$ device we would get the same result also, since whether we call the direction the z-direction or the x-direction cannot affect the results of the experiment.

Putting this all together we can say that

$$|+\hat{\mathbf{y}}\rangle = \langle+\hat{\mathbf{z}}|+\hat{\mathbf{y}}\rangle|+\hat{\mathbf{z}}\rangle + \langle-\hat{\mathbf{z}}|+\hat{\mathbf{y}}\rangle|-\hat{\mathbf{z}}\rangle$$

with

$$|\langle+\hat{\mathbf{z}}|+\hat{\mathbf{y}}\rangle|^2 = |\langle-\hat{\mathbf{z}}|+\hat{\mathbf{y}}\rangle|^2 = \frac{1}{2}$$

and

$$|+\hat{\mathbf{y}}\rangle = \langle+\hat{\mathbf{x}}|+\hat{\mathbf{y}}\rangle|+\hat{\mathbf{x}}\rangle + \langle-\hat{\mathbf{x}}|+\hat{\mathbf{y}}\rangle|-\hat{\mathbf{x}}\rangle$$

with

$$|\langle+\hat{\mathbf{x}}|+\hat{\mathbf{y}}\rangle|^2 = |\langle-\hat{\mathbf{x}}|+\hat{\mathbf{y}}\rangle|^2 = \frac{1}{2}$$

The conventional choice for the phases factors is such that we have

$$|+\hat{\mathbf{x}}\rangle = \frac{1}{\sqrt{2}}|+\hat{\mathbf{z}}\rangle + \frac{1}{\sqrt{2}}|-\hat{\mathbf{z}}\rangle$$

$$|+\hat{\mathbf{y}}\rangle = \frac{1}{\sqrt{2}}|+\hat{\mathbf{z}}\rangle + \frac{i}{\sqrt{2}}|-\hat{\mathbf{z}}\rangle$$

We will derive all of these results more rigorously later.

Chapter 6

Schrodinger Wave equation

1-Dimensional Quantum Systems

The Coordinate Representation

To form a *representation* of an abstract linear vector space we must carry out these steps:

1. Choose a complete, orthonormal set of basis vectors $\{|\alpha_k\rangle\}$.
2. Construct the identity operator $\hat{I}$ as a sum over the one-dimensional subspace projection operators $|\alpha_k\rangle\langle\alpha_k|$

$$\hat{I} = \sum_k |\alpha_k\rangle\langle\alpha_k|$$

3. Write an arbitrary vector $|\psi\rangle$ as a linear combination or superposition of basis vectors using the identity operator

$$|\psi\rangle = \hat{I}|\psi\rangle = \left(\sum_k |\alpha_k\rangle\langle\alpha_k|\right)|\psi\rangle = \sum_k \langle\alpha_k|\psi\rangle|\alpha_k\rangle$$

It is clear from this last equation, that knowledge about the behavior(say in time) of the expansion coefficients $\langle\alpha_k|\psi\rangle$ will tell us the behavior of the state vector $|\psi\rangle$ and allow us to make predictions. Remember also, that the expansion coefficient is the probability amplitude for a particle in the state $|\psi\rangle$ to behave like it is in the state $|\alpha_k\rangle$.

A particular representation that has become very important in the study of many systems using Quantum Mechanics is formed using the eigenstates of the position operator as a basis.

It is called the *coordinate or position representation*.

The eigenstates $\{|\vec{x}\rangle\}$ of the position operator $(\hat{x}, \hat{y}, \hat{z}) = \hat{\mathbf{Q}}$ satisfy

$$\hat{x}\,|\vec{x}\rangle = x\,|\vec{x}\rangle \quad , \quad \hat{y}\,|\vec{x}\rangle = y\,|\vec{x}\rangle \quad , \quad \hat{z}\,|\vec{x}\rangle = z\,|\vec{x}\rangle$$

where the eigenvalues (x, y, z) are continuous variables in the range $[-\infty,\infty]$.

They form the basis of the coordinate representation.

As we mentioned earlier(all the mathematical details are in the book), in this case, all summations in earlier lectures now become integrals in the continuous spectrum case and we have

$$\hat{I} = \int |\vec{x}\rangle\langle\vec{x}|\, d\vec{x} \qquad , \qquad \text{where } d\vec{x} = dxdydz$$

$$|\psi\rangle = \hat{I}\,|\psi\rangle = \int (|\vec{x}\rangle\langle\vec{x}|)\,|\psi\rangle\, d\vec{x} = \int \langle\vec{x}\,|\,\psi\rangle\,|\vec{x}\rangle\, d\vec{x}$$

The expansion coefficient in the coordinate representation is given by

$$\psi(\vec{x}) = \langle\vec{x}\,|\,\psi\rangle$$

Since the inner product is defined for all states $|\vec{x}\rangle$, this new object is clearly a function of the eigenvalues (x, y, z).

As we will see, it is the probability amplitude for finding the particle to be in the neighborhood of the point $\vec{x}$ in 3-dimensional space if it is in the state $|\psi\rangle$.

It is called the **wave function**.

The bra vector or linear functional corresponding to $|\psi\rangle$ is

$$\langle\psi| = \langle\psi|\,\hat{I} = \int \langle\psi|\,(|\vec{x}\rangle\langle\vec{x}|)\, d\vec{x} = \int \langle\psi\,|\,\vec{x}\rangle\langle\vec{x}|\, d\vec{x} = \int \langle\vec{x}\,|\,\psi\rangle^*\langle\vec{x}|\, d\vec{x}$$

The normalization condition takes the form

$$\langle\psi\,|\,\psi\rangle = 1 = \langle\psi|\,\hat{I}\,|\psi\rangle = \int \langle\psi\,|\,\vec{x}\rangle\,\langle\vec{x}\,|\,\psi\rangle\; d\vec{x}$$

$$= \int |\,\langle\vec{x}\,|\,\psi\rangle\,|^2\, d\vec{x} = \int |\psi(\vec{x})|^2\, d\vec{x}$$

$$= \int \psi^*(\vec{x})\psi(\vec{x})\, d\vec{x}$$

The probability amplitude for a particle in the state $|\psi\rangle$ to behave like it is in the state $|\phi\rangle$, where

$$|\phi\rangle = \hat{I}\,|\phi\rangle = \int (|\vec{x}\rangle\,\langle\vec{x}|)\,|\phi\rangle\; d\vec{x} = \int \langle\vec{x}\,|\,\phi\rangle\,|\vec{x}\rangle\; d\vec{x}$$

is given by

$$\langle\phi\,|\,\psi\rangle = \left(\int \langle\vec{x}\,|\,\phi\rangle^*\,\langle\vec{x}|\; d\vec{x}\right)\left(\int \langle\vec{x}\,'\,|\,\psi\rangle\,|\vec{x}\,'\rangle\; d\vec{x}\,'\right)$$

$$= \int d\vec{x} \int d\vec{x}\,'\,\langle\vec{x}\,|\,\phi\rangle^*\,\langle\vec{x}\,'\,|\,\psi\rangle\,\langle\vec{x}\,|\,\vec{x}\,'\rangle$$

We assume the *normalization* condition

$$\langle\vec{x}\,|\,\vec{x}\,'\rangle = \delta(\vec{x} - \vec{x}\,')$$

(position eigenvectors are *not* normalizable).

This normalization condition actually follows (we did not need to assume it) from the expansion in basis states.

We have

$$|\psi\rangle = \int \langle \vec{x}\,' \,|\, \psi \rangle \, |\vec{x}\,'\rangle \, d\vec{x}\,'$$

$$\langle \vec{x} \,|\, \psi \rangle = \int \langle \vec{x}\,' \,|\, \psi \rangle \, \langle \vec{x} \,|\, \vec{x}\,' \rangle \, d\vec{x}\,'$$

$$\psi(\vec{x}) = \int \psi(\vec{x}\,') \, \langle \vec{x} \,|\, \vec{x}\,' \rangle \, d\vec{x}\,'$$

which implies the delta-function normalization.

Thus, the delta function normalization follows from the completeness property of the projection operators or vice versa.

Using this result we get

$$\langle \phi \,|\, \psi \rangle = \int d\vec{x} \int d\vec{x}\,' \, \langle \vec{x} \,|\, \phi \rangle^* \, \langle \vec{x}\,' \,|\, \psi \rangle \, \delta(\vec{x} - \vec{x}\,')$$

$$= \int \langle \vec{x} \,|\, \phi \rangle^* \, \langle \vec{x} \,|\, \psi \rangle \, d\vec{x} = \int \phi^*(\vec{x}) \psi(\vec{x}) \, d\vec{x}$$

We formally write the $\hat{\mathbf{Q}}$ operator using the expansion in eigenvalues and projection operators as

$$\hat{\mathbf{Q}} = \int \vec{x} \, |\vec{x}\rangle \, \langle \vec{x}| \, d\vec{x}$$

This represents three integrals, one for each coordinate.

We will also need the properties of the linear momentum operator.

The eigenstates $\{|\vec{p}\rangle\}$ of the momentum operator ($\hat{p}_x, \hat{p}_y, \hat{p}_z$) = $\hat{\mathbf{P}}$ satisfy

$$\hat{p}_x \, |\vec{p}\rangle = p_x \, |\vec{p}\rangle \quad , \quad \hat{p}_y \, |\vec{p}\rangle = p_y \, |\vec{p}\rangle \quad , \quad \hat{p}_z \, |\vec{p}\rangle = p_z \, |\vec{p}\rangle$$

where the eigenvalues (p_x, p_y, p_z) are continuous variables in the range $[-\infty, \infty]$.

They form the basis of the *momentum representation*.

We then have

$$\hat{I} = \frac{1}{(2\pi\hbar)^3} \int |\vec{p}\rangle \langle \vec{p}| \, d\vec{p}$$

$$|\psi\rangle = \hat{I}\,|\psi\rangle = \frac{1}{(2\pi\hbar)^3} \int (|\vec{p}\rangle \langle \vec{p}|)\,|\psi\rangle \, d\vec{p} = \frac{1}{(2\pi\hbar)^3} \int \langle \vec{p}\,|\,\psi\rangle \, |\vec{p}\rangle \, d\vec{p}$$

The expansion coefficient in the momentum representation is

$$\Psi(\vec{p}) = \langle \vec{p}\,|\,\psi\rangle$$

It is the probability amplitude for finding the particle with momentum $\vec{p}$ (in the neighborhood of) if it is in the state $|\psi\rangle$.

The bra vector or linear functional corresponding to $|\psi\rangle$ is

$$\langle\psi| = \langle\psi|\,\hat{I} = \frac{1}{(2\pi\hbar)^3} \int \langle\psi|\,(|\vec{p}\rangle \langle \vec{p}|)\, d\vec{p} = \frac{1}{(2\pi\hbar)^3} \int \langle \vec{p}\,|\,\psi\rangle^* \langle \vec{p}| \, d\vec{p}$$

The normalization condition takes the form

$$\begin{aligned}\langle \psi\,|\,\psi\rangle = 1 = \langle\psi|\,\hat{I}\,|\psi\rangle &= \frac{1}{(2\pi\hbar)^3} \int \langle \psi\,|\,\vec{p}\rangle \langle \vec{p}\,|\,\psi\rangle \, d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3} \int |\langle \vec{p}\,|\,\psi\rangle|^2 \, d\vec{p} \frac{1}{(2\pi\hbar)^3} = \int |\Psi(\vec{p})|^2 \, d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3} \int \Psi^*(\vec{p})\Psi(\vec{p}) \, d\vec{p}\end{aligned}$$

The probability amplitude for a particle in the state $|\psi\rangle$ to behave like it is in the state $|\phi\rangle$, where

$$|\phi\rangle = \hat{I}\,|\phi\rangle = \frac{1}{(2\pi\hbar)^3} \int (|\vec{p}\rangle \langle \vec{p}|)\,|\phi\rangle \, d\vec{p} = \frac{1}{(2\pi\hbar)^3} \int \langle \vec{p}\,|\,\phi\rangle \, |\vec{p}\rangle \, d\vec{p}$$

is given by

$$\langle\phi|\psi\rangle = \left(\frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}|\phi\rangle^* \langle\vec{p}|\, d\vec{p}\right)\left(\frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,'|\psi\rangle |\vec{p}\,'\rangle\, d\vec{p}\,'\right)$$

$$= \frac{1}{(2\pi\hbar)^6}\int d\vec{p}\int d\vec{p}\,' \langle\vec{p}|\phi\rangle^* \langle\vec{p}\,'|\psi\rangle \langle\vec{p}|\vec{p}\,'\rangle$$

The normalization condition follows from

$$|\psi\rangle = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,'|\psi\rangle |\vec{p}\,'\rangle\, d\vec{p}\,'$$

$$\langle\vec{p}|\psi\rangle = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,'|\psi\rangle \langle\vec{p}|\vec{p}\,'\rangle\, d\vec{p}\,'$$

$$\Psi(\vec{x}) = \frac{1}{(2\pi\hbar)^3}\int \Psi(\vec{p}\,') \langle\vec{p}|\vec{p}\,'\rangle\, d\vec{p}\,'$$

which implies that

$$\frac{1}{(2\pi\hbar)^3}\langle\vec{p}|\vec{p}\,'\rangle = \delta(\vec{p}-\vec{p}\,')$$

Using this result we get

$$\langle\phi|\psi\rangle = \frac{1}{(2\pi\hbar)^6}\int d\vec{p}\int d\vec{p}\,' \langle\vec{p}|\phi\rangle^* \langle\vec{p}\,'|\psi\rangle \langle\vec{p}|\vec{p}\,'\rangle$$

$$= \frac{1}{(2\pi\hbar)^3}\int d\vec{p}\int d\vec{p}\,' \langle\vec{p}|\phi\rangle^* \langle\vec{p}\,'|\psi\rangle \delta(\vec{p}-\vec{p}\,')$$

$$= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}|\phi\rangle^* \langle\vec{p}\psi\rangle\, d\vec{p} = \frac{1}{(2\pi\hbar)^3}\int \Phi^*(\vec{p})\Psi(\vec{p})\, d\vec{p}$$

We formally write the $\hat{\mathbf{P}}$ operator using the expansion in eigenvalues and projection operators as

$$\hat{\mathbf{P}} = \int \vec{p}\,|\vec{p}\rangle\,\langle\vec{p}|\;d\vec{p}$$

We will now derive the *connections* between the two representations.

We have

$$\psi(\vec{x}) = \langle\vec{x}\,|\,\psi\rangle = \langle\vec{x}|\,\hat{I}\,|\psi\rangle = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{x}\,|\,\vec{p}\rangle\,\langle\vec{p}\,|\,\psi\rangle\;d\vec{p}$$

$$= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{x}\,|\,\vec{p}\rangle\,\Psi(\vec{p})\,d\vec{p}$$

and

$$\Psi(\vec{p}) = \langle\vec{p}\,|\,\psi\rangle = \langle\vec{p}|\,\hat{I}\,|\psi\rangle = \int \langle\vec{p}\,|\,\vec{x}\rangle\,\langle\vec{x}\,|\,\psi\rangle\;d\vec{x}$$

$$= \int \langle\vec{x}\,|\,\vec{p}\rangle^*\,\psi(\vec{x})\,d\vec{x}$$

Fourier transforms are defined by

$$g(\vec{p}) = \int e^{-i\vec{p}\cdot\vec{x}/\hbar} f(\vec{x})\,d\vec{x} \quad , \quad f(\vec{x}) = \frac{1}{(2\pi\hbar)^3}\int e^{i\vec{p}\cdot\vec{x}/\hbar} g(\vec{p})\,d\vec{p}$$

which gives us the result

$$\langle\vec{x}\,|\,\vec{p}\rangle = e^{i\vec{p}\cdot\vec{x}/\hbar}$$

It is not a unique choice, however.

It is the choice, however, that allows Quantum mechanics *to make predictions that agree with experiment*.

It is Nature's choice!

We might even say that this choice is *another postulate*.

Now, we can use these results to determine the expectation values of operators involving the position and momentum operators.

Since we are interested in the coordinate representation we need only determine the following quantities.

The position operator calculations are straightforward

$$\langle \vec{x}| \hat{\mathbf{Q}} |\psi\rangle = \vec{x} \langle \vec{x} | \psi\rangle \text{ and } \langle \vec{x}| f(\hat{\mathbf{Q}}) |\psi\rangle = f(\vec{x}) \langle \vec{x} | \psi\rangle$$

For the momentum operator we write

$$\begin{aligned}\langle \vec{x}| \hat{\mathbf{P}} |\psi\rangle &= \frac{1}{(2\pi\hbar)^3} \int \langle \vec{x}| \hat{\mathbf{P}} |\vec{p}\rangle \langle \vec{p} | \psi\rangle \, d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3} \int \langle \vec{x}| \vec{p} |\vec{p}\rangle \langle \vec{p} | \psi\rangle \, d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3} \int \vec{p} \langle \vec{x} | \vec{p}\rangle \langle \vec{p} | \psi\rangle \, d\vec{p}\end{aligned}$$

Now using

$$\langle \vec{x} | \vec{p}\rangle = e^{i\vec{p}\cdot\vec{x}/\hbar}$$

we have

$$\vec{p} \langle \vec{x} | \vec{p}\rangle = -i\hbar\nabla \langle \vec{x} | \vec{p}\rangle = \langle \vec{x}| \hat{\mathbf{P}} |\vec{p}\rangle$$

and thus

$$\begin{aligned}\langle \vec{x}| \hat{\mathbf{P}} |\psi\rangle &= \frac{1}{(2\pi\hbar)^3} \int \vec{p} \langle \vec{x} | \vec{p}\rangle \langle \vec{p} | \psi\rangle \, d\vec{p} \\ &= (-i\hbar\nabla) \frac{1}{(2\pi\hbar)^3} \int (-i\hbar\nabla) \langle \vec{x} | \vec{p}\rangle \langle \vec{p} | \psi\rangle \, d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3} \int \langle \vec{x} | \vec{p}\rangle \langle \vec{p} | \psi\rangle \, d\vec{p} \\ &= -i\hbar\nabla \langle \vec{x} | \psi\rangle\end{aligned}$$

In a similar manner, we can also show that

$$\langle \vec{x} | \hat{\mathbf{P}}^2 | \psi \rangle = (-i\hbar\nabla)^2 = -\hbar^2\nabla^2 \langle \vec{x} | \psi \rangle$$

Alternatively, we can show that the gradient is the correct result using the symmetry transformation ideas we developed earlier.

The momentum operator is the generator of displacements in space.

We showed earlier that

$$e^{-i\vec{a}\cdot\hat{\mathbf{P}}/\hbar} |\vec{x}\rangle = |\vec{x} + \vec{a}\rangle$$

Therefore, we have

$$\begin{aligned} \psi(\vec{x} + \vec{a}) = \langle \vec{x} + \vec{a} | \psi \rangle &= \langle \vec{x} | e^{i\vec{a}\cdot\hat{\mathbf{P}}/\hbar} | \psi \rangle \\ &= e^{-\vec{a}\cdot\nabla} \langle \vec{x} | \psi \rangle = e^{-\vec{a}\cdot\nabla} \psi(\vec{x}) \end{aligned}$$

Since

$$e^{-\vec{a}\cdot\nabla} = 1 - \vec{a}\cdot\nabla + \frac{(\vec{a}\cdot\nabla)^2}{2!} - \ldots$$

we have the standard Taylor series for $\psi(\vec{x} + \vec{a})$.

Therefore, the gradient representation of the momentum operator makes sense.

We now use these results to derive the Schrodinger wave equation.

The Schrodinger wave equation is the partial differential equation that corre-sponds to the eigenvector/ eigenvalue equation for the Hamiltonian operator or the energy operator.

The resulting states are the energy eigenstates.

We already saw that energy eigenstates are stationary states and thus have simple time dependence.

This property will allow us to find the time dependence of amplitudes for very complex systems in a straightforward way.

We have

$$\hat{H}\,|\psi_E\rangle = E\,|\psi_E\rangle$$

where E = a number and $\hat{H}$ = the energy operator = (kinetic energy + potential energy) operators, i.e.,

$$\hat{H} = \frac{\hat{\mathbf{P}}^2}{2m} + V(\hat{\mathbf{Q}})$$

We then have

$$\langle \vec{x}|\left(\frac{\hat{\mathbf{P}}^2}{2m} + V(\hat{\mathbf{Q}})\right)|\psi_E\rangle = E\,\langle \vec{x}\,|\,\psi_E\rangle$$

$$\langle \vec{x}|\,\frac{\hat{\mathbf{P}}^2}{2m}\,|\psi_E\rangle + \langle \vec{x}|\,V(\hat{\mathbf{Q}})\,|\psi_E\rangle = E\,\langle \vec{x}\,|\,\psi_E\rangle$$

$$-\frac{\hbar^2}{2m}\nabla^2\,\langle \vec{x}\,|\,\psi_E\rangle + V(\vec{x})\,\langle \vec{x}\,|\,\psi_E\rangle = E\,\langle \vec{x}\,|\,\psi_E\rangle$$

$$-\frac{\hbar^2}{2m}\nabla^2\psi_E(\vec{x}) + V(\vec{x})\psi_E(\vec{x}) = E\psi_E(\vec{x})$$

which is the *time-independent Schrodinger wave equation*.

The quantity

$$\psi_E(\vec{x}) = \langle \vec{x}\,|\,\psi_E\rangle$$

is the wave function or the energy eigenfunction in the position representation corresponding to energy E.

Now the energy eigenfunctions have a simple time dependence, as we can see from the following.

Since

$$\hat{U}(t)\,|\psi_E\rangle = e^{-i\hat{H}t/\hbar}\,|\psi_E\rangle = e^{-iEt/\hbar}\,|\psi_E\rangle$$

we have

$$|\vec{x}\rangle\,\hat{U}(t)\,|\psi_E\rangle = \psi_E(\vec{x},t) = e^{-iEt/\hbar}\,\langle\vec{x}\,|\,\psi_E\rangle$$

$$\psi_E(\vec{x},t) = e^{-iEt/\hbar}\psi_E(\vec{x},0)$$

Therefore,

$$-\frac{\hbar^2}{2m}\nabla^2\psi_E(\vec{x},t) + V(\vec{x})\psi_E(\vec{x},t) = E\psi_E(\vec{x},t)$$

$$-\frac{\hbar^2}{2m}\nabla^2\psi_E(\vec{x},t) + V(\vec{x})\psi_E(\vec{x},t) = i\hbar\frac{\partial}{\partial t}\psi_E(\vec{x},t)$$

which is the *time-dependent Schrodinger wave equation*.

We will have much more to say about these equations and how to use them later on.

For now, however, let us look at these ideas in a couple of different ways to try and get a better understanding of what they mean.

The Free Particle and Wave Packets

Let us assume that the potential energy term in the one-particle Schrodinger equation is equal to zero.

The solution, in this case, is called a *free particle*.

We can easily see the form of the solution as a function of space and time as follows: the Hamiltonian of the system is

$$\hat{H} = \frac{\hat{\mathbf{P}}^2}{2m} \text{ with } \hat{H}\,|\psi_E\rangle = E\,|\psi_E\rangle$$

This means that $[\,\hat{H}, \hat{\mathbf{P}}\,] = 0$ and thus $\hat{H}$ and $\hat{\mathbf{P}}$ share a common set of eigenvectors(eigenfunctions now).

The eigenvectors(eigenfunctions) corresponding to linear momentum $\hat{\mathbf{P}}$ are easy to find from our earlier derivations.

We have

$$\hat{\mathbf{P}}\,|\vec{p}\rangle = \hat{\mathbf{P}}\,|\psi_E\rangle = \vec{p}\,|\vec{p}\rangle = \vec{p}\,|\psi_E\rangle$$

$$\langle\vec{x}|\,\hat{\mathbf{P}}\,|\psi_E\rangle = \vec{p}\,\langle\vec{x}\,|\,\psi_E\rangle = \vec{p}e^{i\vec{p}\cdot\vec{x}/\hbar}$$

$$= -i\hbar\nabla e^{i\vec{p}\cdot\vec{x}/\hbar} = -i\hbar\nabla\,\langle\vec{x}\,|\,\psi_E\rangle$$

which says that the eigenfunctions are

$$\langle\vec{x}\,|\,\vec{p}\rangle = e^{i\vec{p}\cdot\vec{x}/\hbar} = \psi_E(\vec{x}) = \langle\vec{x}\,|\,\psi_E\rangle = \langle\vec{x}\,|\,E\rangle$$

We can see that this satisfies the time-independent Schrodinger equation with no potential term.

This solution is also called a *plane-wave*.

If we assume that at time $t = 0$, the state of the particle is

$$\psi_E(\vec{x}, 0) = e^{i\vec{p}\cdot\vec{x}/\hbar}$$

then the time dependence is given by

$$|\vec{x}, t\rangle = \hat{U}(t)\,|\vec{x}, 0\rangle = e^{-iEt/\hbar}\,|\vec{x}, 0\rangle$$

$$\langle\vec{x}, t\,|\,\psi_E\rangle = e^{-iEt/\hbar}\,\langle\vec{x}, 0\,|\,\psi_E\rangle$$

$$\psi_E(\vec{x}, t) = e^{-iEt/\hbar}\psi_E(\vec{x}, 0) = e^{-iEt/\hbar}e^{i\vec{p}\cdot\vec{x}/\hbar} = e^{i(\vec{p}\cdot\vec{x}-Et)/\hbar}$$

where $E = p^2/2m$.

Now we can write the position state vector the as a linear combination of momentum state vectors or correspondingly we can write position space wave function as a linear combination of momentum eigenfunctions or plane waves (since they are a basis).

$$|\psi_E\rangle = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\psi_E\rangle\,|\vec{p}\rangle\; d\vec{p}$$

$$\langle\vec{x}\,|\,\psi_E\rangle = \psi_E(\vec{x}) = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\psi_E\rangle\,\langle\vec{x}\,|\,\vec{p}\rangle\; d\vec{p}$$

$$\psi_E(\vec{x},0) = \frac{1}{(2\pi\hbar)^3}\int e^{i\vec{p}\cdot\vec{x}/\hbar}\Psi_E(\vec{p},0)\, d\vec{p}$$

We note that in the state $|\vec{p}\rangle$, which is an eigenvector of $\hat{\mathbf{P}}$, the momentum has the value $\vec{p}$ with probability = 1 (a certainty), which is the meaning of an eigenvector!

However, the probability of finding the particle in this state at the point $\vec{x}$ is independent of $\vec{x}$, i.e.,

$$|\langle\vec{x}\,|\,\vec{p}\rangle|^2 = |e^{i\vec{p}\cdot\vec{x}/\hbar}|^2 = 1$$

This means that the particle is equally probable to have any value of momentum.

The momentum is completely uncertain in a position eigenstate.

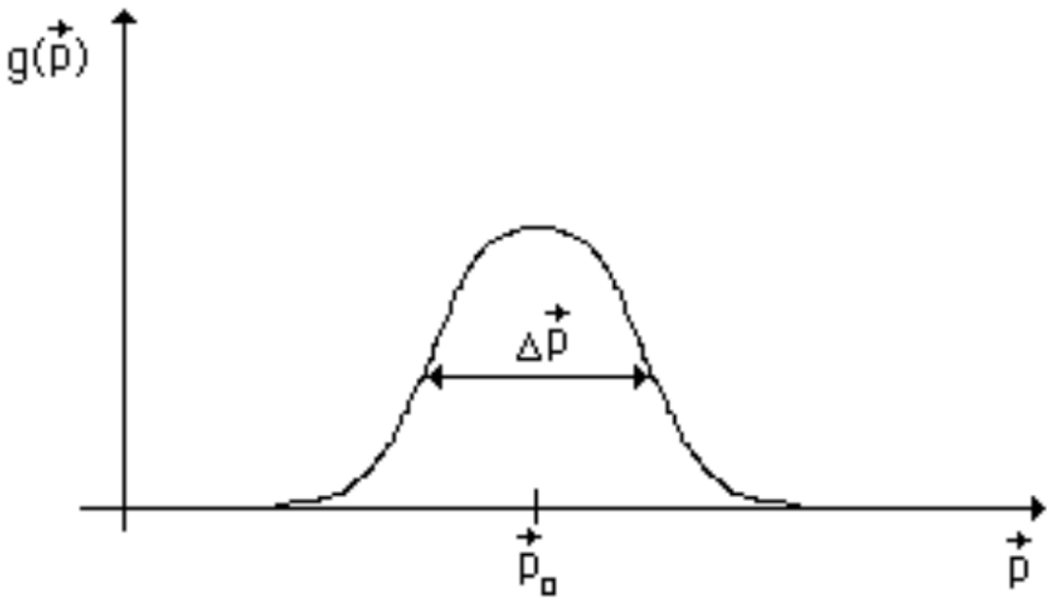
If we consider an intermediate case in which the particle is reasonably well localized in position space, and at the same time, has a fairly well defined momentum, then the wave function for such a state is called a *wave packet*.

Suppose that the function $|\Psi_E(\vec{p},0)|$ is peaked about the value $\vec{p}_0$, with a width $\approx \Delta\vec{p}$.

In particular, at $t = 0$, we choose

$$\Psi_E(\vec{p},0) = g(\vec{p})e^{i\alpha(\vec{p})}$$

where $g(\vec{p})$, called the weight function, takes the form of a real, nonnegative function such as shown in in the Figure, and the phase $\alpha(\vec{p})$ is real.



Weight Function

We then have

$$\psi_E(\vec{x},t) = \frac{1}{(2\pi\hbar)^3}\int g(\vec{p})e^{i(\vec{p}\cdot\vec{x}-Et+\hbar\alpha(\vec{p}))/\hbar}\,d\vec{p}$$

as the wave function of a particle whose momentum is approximately $\vec{p}_0$.

This means that most of the time, if we measure the momentum of the particle in this state, then we would find a value within $\Delta\vec{p}$ of $\vec{p}_0$.

We now determine how this particle moves in time and why we have called this expression a wave packet.

If for some position $\vec{x}$ (which is a parameter in the integral) the phase of the exponential term is *not slowly varying* for some values of integration variable $\vec{p}$, then the integral will equal zero.

This occurs because the exponential term is a very rapidly oscillating function (since $\hbar$ is very small), i.e., if you multiply a very rapidly varying function by a function of the assumed form of $g(\vec{p})$, then *on the averag*e the integrand is zero.

In effect, we are seeing destructive interference between the plane waves.

If, however, $\vec{x}$ is a point such that the phase of the exponential remains fairly constant over the same range of $\vec{p}$ where the function $g(\vec{p})$ is nonzero, then we will get a nonzero value for the integral and the particle will have some nonzero probability of being at that point.

One way to make this work is as follows:

1. At time t, assume the point $\vec{x}_t$ makes the phase *stationary* near $\vec{p}_0$. By stationary we mean

$$\nabla_{\vec{p}}(\vec{p}\cdot\vec{x}_t - Et + \hbar\alpha(\vec{p}))_{\vec{p}=\vec{p}_0} = 0$$

   or we are near an extremum of the phase and hence the phase will be slowly varying.

2. We will then get strong *constructive* interference between the plane waves and $\psi_E(\vec{x},t)$ will be nonzero.

Solving the stationary phase equation for $\vec{x}_t$ we find, using $E = p^2/2m$

$$(\vec{x}_t - \nabla_{\vec{p}}(E)t + \hbar(\nabla_{\vec{p}}\alpha(\vec{p})))_{\vec{p}=\vec{p}_0} = 0$$

$$\vec{x}_t = ((\vec{p}E)t + \hbar(\nabla_{\vec{p}}\alpha(\vec{p})))_{\vec{p}=\vec{p}_0}$$

$$= \frac{\vec{p}_0}{m}t - \hbar(\nabla_{\vec{p}}\alpha(\vec{p})) = \frac{\vec{p}_0}{m}t + \vec{x}_0$$

Therefore, the point $\vec{x}_t$, which is essentially the center of the wave packet, moves in time with a constant velocity $\vec{v} = \vec{p}_0/m$.

This is exactly the velocity one expects for a free particle of momentum $\vec{p}_0$ and mass m.

It is called the *group velocity* of the wave packet.

This procedure is called the *stationary phase method*.

How localized is this wave packet(*particle*) or what is the spatial extent of the packet?

We will have a nonzero integral and hence a nonzero probability for the particle to be at a particular point as long as we have constructive interference or as long as the exponential undergoes less than one oscillation as $\vec{p}$ varies over the region for which the function $g(\vec{p})$ is also large.

The change in the phase of the exponential as we vary the x-component of $\vec{p}$ is approximately

$$\Delta\phi = \Delta(\vec{p}\cdot\vec{x} - Et + \hbar\alpha(\vec{p}))_{\vec{p}=\vec{p}_0} = \frac{1}{\hbar}\Delta p_x \frac{\partial}{\partial p_x}(\vec{p}\cdot\vec{x} - Et + \hbar\alpha(\vec{p}))_{\vec{p}=\vec{p}_0}$$

$$= \frac{1}{\hbar}\Delta p_x \left(x - \frac{\partial}{\partial p_x}(Et + \hbar\alpha(\vec{p})\right)_{\vec{p}=\vec{p}_0} = \frac{1}{\hbar}\Delta p_x (x - x_t)$$

As long as $\Delta\phi \leq 2\pi$ we will get constructive interference and the integral will be nonzero.

This tells us the extent of the wave packet in the x-direction about the point x_t.

We get

$$|x - x_t| \geq \frac{2\pi\hbar}{\Delta p_x} \text{ or } |x - x_t|\Delta p_x \geq h$$

or

$$\Delta x \Delta p_x \geq h$$

Similarly we can show that $\Delta y \Delta p_y \geq h$ and $\Delta z \Delta p_z \geq h$.

These relations between the spatial extent of the wave packet or the uncertainty in its position, and the uncertainty in the particle momentum, are often called the *Heisenberg uncertainty principle*.

We will see later that the uncertainty principle involves more than just these simple properties of a wave packet.

We will also develop more details about wave packets later on thlectures.

Note that there is no uncertainty relation between

$$\Delta x \text{ and } \Delta y, \Delta z, \Delta p_y, \Delta p_z$$
$$\Delta y \text{ and } \Delta x, \Delta z, \Delta p_x, \Delta p_z$$
$$\Delta z \text{ and } \Delta x, \Delta y, \Delta p_x, \Delta p_y$$

What is the difference between these pairs of variables?

The difference is in their commutators.

It turns out, in this example of the wave packet, that only those observables that are represented by non-commuting operators have nonzero uncertainty relations.

This result has nothing to do with wave packets, Schrodinger equations, wave functions or Fourier transforms, although all of these quantities can be used to show that uncertainty relations exist.

The existence of uncertainty relations between non-commuting observables was actually built into the theory of Quantum Mechanics when we assumed that we were in an abstract linear vector space.

All linear vector spaces have a property called the Schwarz inequality and it always leads to such uncertainty relations for non-commuting operators.

So the Heisenberg uncertainty principle, to which much mysticism has been attached, is really an assumption and little more than a lemma that follows from the original assumption of a vector space.

Derivation of the Uncertainty Relations in General

Given two Hermitian operators $\hat{A}$ and $\hat{B}$ we define the two new operators

$$\hat{D}_A = \hat{A} - \langle \hat{A} \rangle \text{ and } \hat{D}_B = \hat{B} - \langle \hat{B} \rangle$$

where $\langle \hat{O} \rangle = \langle \psi | \hat{O} | \psi \rangle$ equals the average or expectation value in the state $|\psi\rangle$.

In the statistical analysis of data, we use a quantity called the standard or mean-square deviation as a measure of the uncertainty of an observed quantity.

It is defined, for a set of *N* measurements of the quantity Δq by

$$(\Delta q)^2 = (\text{standard deviation})^2 = \frac{1}{N}\sum_{i=1}^{N}(q_i - q_{average})^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}(q_i)^2 - \frac{1}{N}\sum_{i=1}^{N}(q_i q_{average}) - \frac{1}{N}\sum_{i=1}^{N}(q_{average} q_i) + \frac{1}{N}\sum_{i=1}^{N}(q_{average})^2$$

$$= (q^2)_{average} - (q_{average})^2$$

where we have used

$$q_{average} = \frac{1}{N}\sum_{i=1}^{N} q_i$$

In analogy, we define the mean-square deviations for $\hat{A}$ and $\hat{B}$ as

$$(\Delta\hat{A})^2 = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2 = \langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle = \langle\hat{D}_A^2\rangle$$
$$(\Delta\hat{B})^2 = \langle\hat{B}^2\rangle - \langle\hat{B}\rangle^2 = \langle(\hat{B} - \langle\hat{B}\rangle)^2\rangle = \langle\hat{D}_B^2\rangle$$

We then have

$$(\Delta\hat{A})^2(\Delta\hat{B})^2 = \langle\hat{D}_A^2\rangle\langle\hat{D}_B^2\rangle$$

Now we assume that

$$[\hat{D}_A, \hat{B}] = [\hat{A}, \hat{B}] = [\hat{D}_A, \hat{D}_B] = i\hat{C}$$

where $\hat{C}$ is also a Hermitian operator and we let

$$|\alpha\rangle = \hat{D}_A|\psi\rangle = (\hat{A} - \langle\hat{A}\rangle)|\psi\rangle \text{ and } |\beta\rangle = \hat{D}_B|\psi\rangle = (\hat{B} - \langle\hat{B}\rangle)|\psi\rangle$$

Then we have

$$(\Delta\hat{A})^2 = \langle\hat{D}_A^2\rangle = \langle\psi|\hat{D}_A^2|\psi\rangle = (\langle\psi|\hat{D}_A)(\hat{D}_A|\psi\rangle) = \langle\alpha|\alpha\rangle$$
$$(\Delta\hat{B})^2 = \langle\hat{D}_B^2\rangle = \langle\psi|\hat{D}_B^2|\psi\rangle = (\langle\psi|\hat{D}_B)(\hat{D}_B|\psi\rangle) = \langle\beta|\beta\rangle$$

The Schwarz inequality says that for any two vectors we must have the relation

$$\langle\alpha|\alpha\rangle\langle\beta|\beta\rangle \geq |\langle\alpha|\beta\rangle|^2$$

We therefore have

$$(\Delta\hat{A})^2(\Delta\hat{B})^2 = \langle\hat{D}_A^2\rangle\langle\hat{D}_B^2\rangle = \langle\alpha|\alpha\rangle\langle\beta|\beta\rangle \geq |\langle\alpha|\beta\rangle|^2$$
$$(\Delta\hat{A})^2(\Delta\hat{B})^2 \geq |\langle\psi|\hat{D}_A\hat{D}_B|\psi\rangle|^2 = |\langle\hat{D}_A\hat{D}_B\rangle|^2$$

Now

$$|\langle \hat{D}_A \hat{D}_B \rangle|^2 = |\langle \Delta\hat{A}\Delta\hat{B} \rangle|^2 = |\langle \frac{1}{2}[\Delta\hat{A}, \Delta\hat{B}] + \frac{1}{2}\{\Delta\hat{A}, \Delta\hat{B}\} \rangle|^2$$

$$= |\langle \frac{1}{2}[\hat{A}, \hat{B}] + \frac{1}{2}\{\hat{A}, \hat{B}\} \rangle|^2$$

where

$$[\Delta\hat{A}, \Delta\hat{B}]^\dagger = -[\Delta\hat{A}, \Delta\hat{B}] \rightarrow \text{ anti-Hermitian } \rightarrow \text{ expectation value is imaginary}$$

$$\{\Delta\hat{A}, \Delta\hat{B}\}^\dagger = \{\Delta\hat{A}, \Delta\hat{B}\} \rightarrow \text{ Hermitian } \rightarrow \text{ expectation value is real}$$

Therefore,

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

$$\geq \frac{1}{4}|\langle i\hat{C} \rangle + a|^2$$

where a is a real number.

Then

$$(\Delta\hat{A})^2(\Delta\hat{B})^2 \geq \frac{1}{4}|a|^2 + \frac{1}{4}|\langle \hat{C} \rangle|^2 \geq \frac{1}{4}|\langle \hat{C} \rangle|^2$$

since $|a|^2/4 \geq 0$.

This is the Heisenberg uncertainty principle.

It is simply the Schwarz inequality!!

The standard form in most texts follows from $[\ \hat{x}, \hat{p}_x\ ] = \mathrm{i}\,\hbar$, which gives $\Delta\hat{x}\Delta\hat{p}_x \geq \frac{\hbar}{2}$

We note that if $[\ \hat{A}, \hat{B}\ ] = \mathrm{i}\,\hat{C} = 0$, we then get

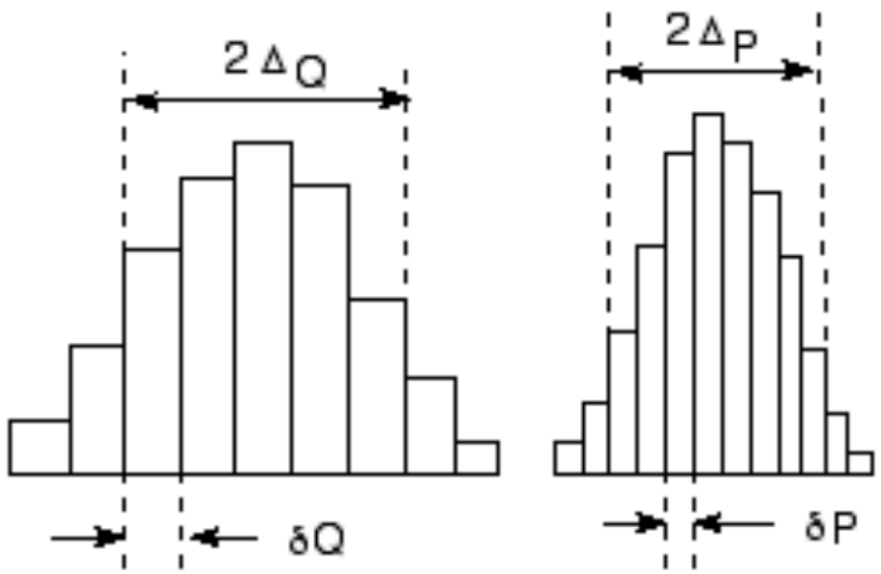
$$(\Delta\hat{A})^2(\Delta\hat{B})^2 \geq 0$$

or commuting observables do not have an uncertainty principle!

The Meaning of the Indeterminacy Relations

What is the significance of indeterminacy relations in the world of experimental physics?

Consider the experimental results shown in the Figure below:



Experimental Data

These are frequency distributions for the results of independent measurements of Q and P on an ensemble if similarly prepared systems, i.e., on each of a large number of similarly prepared systems one performs a single measurement (either Q or P).

The histograms are the statistical distribution of the results.

The standard deviations (variances) as shown must satisfy (according to the theory) the relation

$$\Delta_Q \Delta_P \geq \frac{\hbar}{2}$$

They must be distinguished from the resolution of the individual measurements, δQ and δP.

Let me emphasize these points:

1. The quantities Δ_Q and Δ_P are *not* errors of measurement. The *errors* or preferably the *resolutions* of the Q and P measuring instruments are δ_Q and δ_P. They are logically unrelated to Δ_Q and Δ_P and to the uncertainty relations except for the practical requirement that if

$$\delta_Q > \Delta_Q \ (or\ \delta_P > \Delta_P) \tag{6.71}$$

   then it will not be possible to determine Δ_Q (or Δ_P) in the experiment and the experiment cannot test the uncertainty relation.

2. The experimental test of the indeterminacy relation *does not* involve simultaneous measurements of Q and P, but rather it involves the measurement of one or the other of these dynamical variables on each independently prepared representative of the particular state being studied.

Why am I being so picky here?

The quantities ΔQ and ΔP as defined here are often misinterpreted as the errors of individual measurements.

This probably arises because Heisenberg's original paper on this subject, published in 1927, was based on an early version of quantum mechanics that predates the systematic formulation and statistical interpretation of quantum mechanics as it exists now.

The derivation, as carried out here was not possible in 1927!

Time-Energy Uncertainty Relations

The use of time-energy uncertainty relations in most textbooks is simply incorrect.

Let us now derive the most we can say about such relations.

Earlier, we showed that

$$\frac{d\langle \hat{Q} \rangle_t}{dt} = Tr\left(\hat{W}_0 \frac{d\hat{Q}_H(t)}{dt} \right) = Tr\left[\frac{i}{\hbar} \hat{W}_0 [\hat{H}, \hat{Q}_H(t)] + \hat{W}_0 \left(\frac{\partial \hat{Q}}{\partial t} \right)_H \right]$$

in the Heisenberg picture.

So we can write in general that for any operator $\hat{Q}$

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

Now consider a system whose Hamiltonian $\hat{H}$ does not explicitly depend on time and let $\hat{Q}$ be another observable of this system which does not depend on time explicitly so that

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

We consider the dynamical state of the system at a given time *t*.

Let $|\psi\rangle$ be the vector representing that state.

Call ΔQ and ΔE the root-mean-square deviations of and $\hat{H}$, respectively.

Applying the Schwarz inequality (as earlier) to the vectors $(\hat{Q} - \langle \hat{Q} \rangle)|\psi\rangle$ and $(\hat{H} - \langle \hat{H} \rangle)|\psi\rangle$ and carrying out the same manipulations, we find after some calculations

$$\Delta Q \Delta E \geq \frac{1}{2} |\langle [\hat{Q}, \hat{H}] \rangle|$$

the equality being realized when $|\psi\rangle$ satisfies the equation

$$(\hat{Q} - \alpha)\,|\psi\rangle = i\gamma(\hat{H} - \epsilon)\,|\psi\rangle$$

where α, γ, and ε are arbitrary real constants.

We then have

$$\frac{\Delta Q}{|\frac{d\langle \hat{Q} \rangle}{dt}|} \Delta E \geq \frac{\hbar}{2}$$

or

$$\tau_Q \Delta E \geq \frac{\hbar}{2}$$

where we have defined

$$\tau_Q = \frac{\Delta Q}{|\frac{d\langle \hat{Q} \rangle}{dt}|}$$

τ_Q appears as a time characteristic of the evolution of the expectation value of $\hat{Q}$.

It is the time required for the center of the center $\langle \hat{Q} \rangle$ of the statistical distribution of $\hat{Q}$ to be displaced by an amount equal to its width ΔQ.

In other words, the time necessary for this statistical distribution to be appreciably modified.

In this way we can define a characteristic evolution time for each dynamical variable of the system.

Let τ be the shortest of the times thus defined.

τ may be considered as the characteristic time of evolution of the system itself, that is, whatever the measurement carried out on the system at an instant of time t', the statistical distribution of the results is essentially the same as would be obtained at the instant t, as long as the difference $|t - t'|$ is less than τ.

This time τ and the energy spread ΔE satisfy the time-energy uncertainty relation

$$\tau \Delta E \geq \frac{\hbar}{2}$$

If, in particular, the system is in a stationary state where $d\langle \hat{Q} \rangle/dt = 0$ no matter what $\hat{Q}$, and consequently τ is infinite, then $\Delta E = 0$ according to the above relation.

Ordinary time t is just a parameter in non-relativistic QM and not an operator!

We cannot say that

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

which is an equation that has no meaning!

The Wave Function and Its Meaning

The Schrodinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi_E(\vec{x},t) + V(\vec{x})\psi_E(\vec{x},t) = i\hbar\frac{\partial}{\partial t}\psi_E(\vec{x},t)$$

has the mathematical form similar to that of a type of equation called a *wave equation*.

Since other wave equations imply the existence of real physical fields or waves propagating in real three dimensional space, a possible interpretation of the Schrodinger wave function $\psi_{\mathrm{E}}(\vec{\mathrm{x}},t)$ is that of a wave in real, three dimensional space.

We might also associate the wave field with a particle or a particle could be identified with a wave packet solution to the Schrodinger equation, as we did earlier in our discussion.

These are all misinterpretations!!!

We are being misled here because we are working with a simple system, namely, a single particle in three-dimensional space.

To see that these interpretations are not valid, we must look at the more complicated case of a system of N particles.

We now generalize to a coordinate representation for the system of N particles as follows:

1. Choose as a basis the set of vectors that is a common set of eigenvectors for the N position operators $\hat{\mathbf{Q}}^{(1)}, \hat{\mathbf{Q}}^{(2)}, \ldots, \hat{\mathbf{Q}}^{(\mathbf{N})}$ corresponding to the N particles.

2. Assuming that each position operator satisfies an eigenvalue/eigenvector equation of the form

$$\hat{Q}^{(i)} \left|\vec{x}^{(i)}\right\rangle = \vec{x}^{(i)} \left|\vec{x}^{(i)}\right\rangle \quad , \quad i = 1, 2, \ldots, N$$

and that each of the sets of single particle eigenvectors forms a basis for the 3-dimensional subspace of the single particle. The basis states for the N particle system are then the *direct product* states among the N sets of single particle eigenvectors. We write them as

$$\left|\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}\right\rangle = \left|\vec{x}^{(1)}\right\rangle \otimes \left|\vec{x}^{(2)}\right\rangle \otimes \ldots \otimes \left|\vec{x}^{(N)}\right\rangle$$

The state vector $|\Psi\rangle$ representing an N particle system is then represented in the $3N$-dimensional configuration space corresponding to the $3N$ coordinates of the N particles by

$$\Psi(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}) = \left\langle \vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)} \middle| \Psi \right\rangle$$

The Hamiltonian for the N particle system is given by

$$\hat{H} = \sum_{n=1}^{N} \hat{H}_n + \hat{U}(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)})$$

where

$$\hat{H}_n = \text{ single particle Hamiltonian for the } n^{th} \text{ particle}$$

$$= -\frac{\hbar^2}{2m_n}\nabla_n^2 + V(\vec{x}^{(n)})$$

and $\hat{U}(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)})$ = the interparticle interaction potential energy.

The equation of motion(the N-particle Schrodinger equation) is then

$$\hat{H}\Psi(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}) = i\hbar\frac{\partial}{\partial t}\Psi(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)})$$

As we shall now see, this N particle equation does not allow any of the interpretations that might seem to be valid for the single particle Schrodinger equation.

If we associated a physical wave in real, three dimensional space with a particle or if a particle were to be identified with a wave packet, then there would have to be N interacting waves in the real, three dimensional space where we actually measure things.

It is clear, however, that the N particle Schrodinger equation says that this is NOT the case.

There is only one wave function in an abstract $3N$-dimensional configuration space.

We are able to make the incorrect interpretation in the one particle case because it just happens, in that case, that the real, 3-dimensional has a one-to-one correspondence to the 3-dimensional configuration space of the Schrodinger equation.

The proper interpretation of Ψ is that it is a statistical state function.

It is nothing more than a function that enables us to calculate probability distributions for all observables and their expectation values.

The important physical quantities are the probabilities and the NOT the function used to calculate them.

$$|\Psi(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)})|^2 = \text{ probability density in configuration space for}$$

$$\text{particle 1 to be at } \vec{x}^{(1)}$$

$$\text{particle 2 to be at } \vec{x}^{(2)}$$

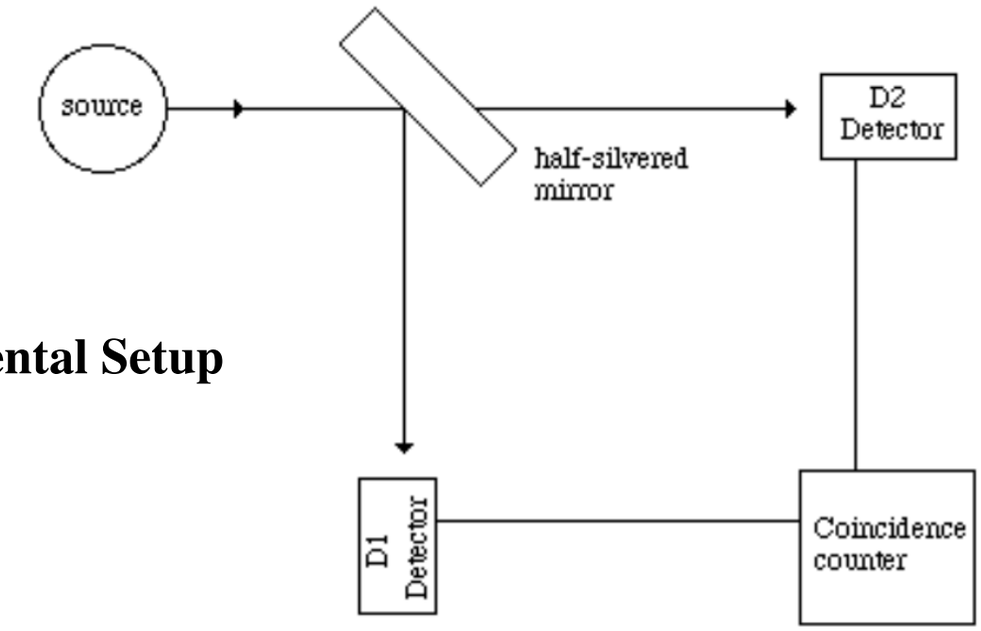
$$\ldots$$

$$\ldots$$

$$\text{particle N to be at } \vec{x}^{(N)}$$

We can demonstrate the necessity for a purely statistical interpretation for the *N*-particle wave function as follows.

Consider the experiment with photons shown in the Figure.



Experimental Setup

We have a source which emits light(photons).

This light falls on a half-silvered mirror, which allows 1/2 of the light to be transmitted though and 1/2 to be reflected as shown.

The transmitted and reflected light is detected with the two photomultipliers $D1$ and $D2$, respectively.

Finally signals generated by the detectors when they record a photon hit, are sent to a coincidence counter, which records a count when it receives two signals at the same time (one from each detector) to within some specified accuracy.

Now, suppose that the source can be described as an emitter of identical localized wave packets ($|\Psi|^2$ is nonzero only over a finite region of space).

Assume that these packets are emitted at random times, at an average rate R per second and that no more than one packet exists in a small interval Δ.

The probability of an emission in any short interval of length Δt is then

$$p = R\Delta t$$

In each of these intervals, the detectors either see a photon or they do not.

The experimental arrangement guarantees that over long periods of time each detector sees 1/2 of the emitted photons.

We can learn more about the details of what is happening as the detectors record photon hits by monitoring the temporal response of the two detectors, in particular, we can ask whether the detectors have both recorded photon hits during the same interval.

The experimental procedure is to send light into the system and record the number of coincidences relative to number of individual counts of the detectors.

The results are analyzed in terms of an anti-coincidence parameter A given by

$$A = \frac{P_c}{P_1 P_2}$$

where

P_1 = experimentally measured probability of detector 1 responding

P_2 = experimentally measured probability of detector 2 responding

P_c = experimentally measured probability of coincidences

If light is composed of single particles (photons) and photons are not divisible and $|\Psi(\vec{x},t)|^2$ = the probability per unit volume at time t that the photon will be located within some small volume about the point $\vec{x}$, then the two detectors should never both record a hit in the same interval (they are mutually exclusive events) and thus $P_c = A = 0$.

As we shall see later on when we analyze the space-time behavior of wave packets, if the photon is a wave packet, then the packet will split into two equal parts, one on each path.

This means that both detectors will always record hits during the same interval and thus $P_c = 1$.

The probability of a detector recording a count is proportional to the amount of the wave amplitude in the region of the detector, in particular, to

$$\int_{\substack{detector \\ volume}} |\Psi|^2 \, d\vec{x}$$

For a symmetrical, equal path device, this is exactly $p/2$ for each detector.

If the detectors are far apart, the respective triggerings are independent (we assume that a spacelike interval exists between the events).

Therefore, the probability of coincidence is given by

$$P_c = \frac{p^2}{4}$$

Experiments of this type were carried out by Clauser(1974) and Grangier, Roger and Aspect(1986).

They obtained $A = 0$.

This confirmed that light is photons or single particles with a statistical interpretation of the wave function and not particles repreented by wave packets or wave fields of some kind.

One-Dimensional Systems

The time-independent Schrodinger equation in 1-dimension is

$$-\frac{\hbar^2}{2m}\nabla^2\psi_E(x) + V(x)\psi_E(x) = E\psi_E(x)$$

The solutions $\psi_{\mathrm{E}}(x)$ are the energy eigenstates (eigenfunctions).

As we have seen, their time dependence is given by

$$\psi_E(x,t) = e^{-iEt/\hbar}\psi_E(x,0) \text{ where } \psi_E(x,0) = \langle x \,|\, E\rangle$$

and

$$\hat{H}\,|E\rangle = E\,|E\rangle \text{ where } \hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x})$$

We are thus faced with solving an ordinary differential equation with boundary conditions.

Since $\psi_{\mathrm{E}}(x)$ is physically related to a probability amplitude and hence to a measurable probability, we assume that $\psi_{\mathrm{E}}(x)$ is continuous.

Using this assumption, we can determine the general continuity properties of $d\psi_{\mathrm{E}}(x)/dx$.

The continuity property at a particular point, say $x = x_0$, is derived as follows:

$$\int_{x_0-\epsilon}^{x_0+\epsilon} \frac{d^2\psi_E(x)}{dx^2}\,dx = \int_{x_0-\epsilon}^{x_0+\epsilon} d\left(\frac{d\psi_E(x)}{dx}\right)$$

$$= -\frac{2m}{\hbar^2}\left[E\int_{x_0-\epsilon}^{x_0+\epsilon}\psi_E(x)\,dx - \int_{x_0-\epsilon}^{x_0+\epsilon} V(x)\psi_E(x)\,dx\right]$$

Taking the limit as $\varepsilon \to 0$ we have

$$\lim_{\epsilon\to 0}\left(\left.\frac{d\psi_E(x)}{dx}\right|_{x=x_0+\epsilon} - \left.\frac{d\psi_E(x)}{dx}\right|_{x=x_0-\epsilon}\right)$$

$$= -\frac{2m}{\hbar^2}\left[E\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon}\psi_E(x)\,dx - \lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon} V(x)\psi_E(x)\,dx\right]$$

or

$$\Delta\left(\frac{d\psi_E(x)}{dx}\right) = \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon} V(x)\psi_E(x)\,dx$$

where we have used the continuity of $\psi_E(x)$ to set

$$\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon}\psi_E(x)\,dx = 0$$

This make it clear that whether or not $d\psi_E(x)/dx$ has a discontinuity depends directly on the properties of the potential energy function.

If $V(x)$ is continuous at $x = x_0$ (harmonic oscillator example later), i.e., if

$$\lim_{\epsilon\to 0}[V(x_0+\epsilon) - V(x_0-\epsilon)] = 0$$

then

$$\Delta\left(\frac{d\psi_E(x)}{dx}\right) = \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon} V(x)\psi_E(x)\,dx = 0$$

and $d\psi_E(x)/dx$ is continuous at $x = x_0$.

If $V(x)$ has a *finite* discontinuity (jump) at $x = x_0$ (finite square well and square barrier examples later), i.e., if

$$\lim_{\epsilon\to 0}[V(x_0+\epsilon) - V(x_0-\epsilon)] = \text{ finite}$$

then

$$\Delta\left(\frac{d\psi_E(x)}{dx}\right) = \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon} V(x)\psi_E(x)\,dx = 0$$

and $d\psi_E(x)/dx$ is continuous at $x = x_0$.

Finally, if $V(x)$ has an infinite jump at $x = x_0$ (infinite square well and delta-function examples later), then we have two choices:

1. if the potential is infinite over an extended range of x (the infinite well), then we must force $\psi_E(x) = 0$ in that region and use only the continuity of $\psi_E(x)$ as a boundary condition at the edge of the region

2. if the potential is infinite at a single point, i.e., $V(x) = A\delta(x - x_0)$, then

$$\begin{aligned}\Delta\left(\frac{d\psi_E(x)}{dx}\right) &= \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon} V(x)\psi_E(x)\,dx \\ &= \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon} A\delta(x-x_0)\psi_E(x)\,dx \\ &= \frac{2mA}{\hbar^2}\lim_{\epsilon\to 0}\psi_E(x_0) = \frac{2mA}{\hbar^2}\psi_E(x_0)\end{aligned}$$

and, thus, $d\psi_E(x)/dx$ is discontinuous at $x = x_0$.

The last thing we must worry about is the validity of our probability interpretation of $\psi_E(x)$, i.e.,

$$\psi_E(x) = \langle x \,|\, \psi_E \rangle = \text{probability amplitude for the particle in the state } |\psi_E\rangle \text{ to be found at } x$$

which says that we must also have

$$\langle \psi_E \,|\, \psi_E \rangle = \int_{-\infty}^{\infty} |\psi_E(x)|^2 \, dx < \infty$$

This means that we must be able to normalize the wave functions and make the total probability that the particle is somewhere on the x-axis equal to one.

A wide range of interesting physical systems can be studied using 1-dimensional potential energy functions.

We will consider potentials in the form of square wells and square barriers, delta-functions, linear functions and parabolic functions.

We will now learn how to work with potentials in 1-dimension by doing a variety of different examples.

Along the way we will learn some very clever techniques and tricks and expand our understanding of the physics and mathematics of Quantum Mechanics.

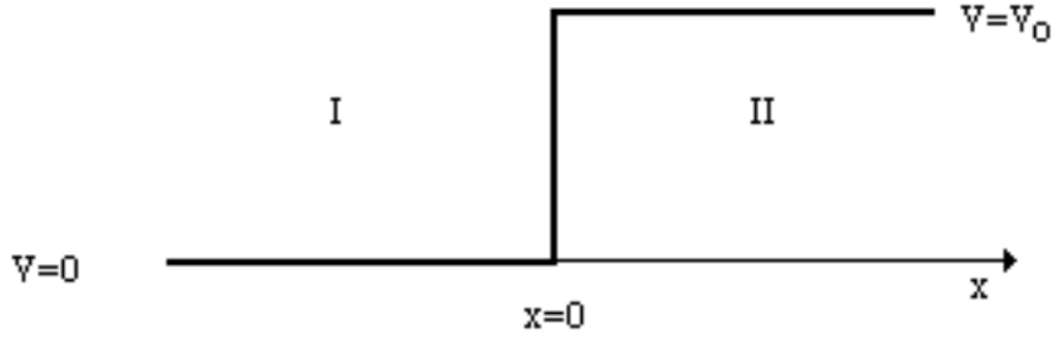
We start with a simple system to illustrate the process for solving the Schrodinger equation.

One-Dimensional Barrier

Consider the potential energy function

$$V(x) = \begin{cases} 0 & x \le 0 \\ V_0 & x \ge 0 \end{cases}$$

which looks like the Figure below.



Finite Step barrier

Since there are no infinite jumps in the potential energy function, both $\psi(x)$ and $d\psi(x)/dx$ are continuous everywhere.

We have two distinct regions to consider, which are labeled I and II in the Figure.

In region I, $V(x) = 0$ and the Schrodinger equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_I(x)}{dx^2} = E\psi_I(x)$$

Let us define k by

$$E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}$$

There are two possible solutions to this equation in region I, namely,

$$\psi_I(x) = e^{\pm ikx}$$

each with energy

$$E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}$$

The most general solution is a linear combination of the two possible solutions

$$\psi_I(x) = Ae^{+ikx} + Be^{-ikx}$$

This is a linear combination of two waves.

Since this is an energy eigenstate it has a time dependence

$$e^{-iEt/\hbar}$$

If we insert the time dependence we have

$$\psi_I(x,t) = Ae^{+i(kx-\omega t)} + Be^{-i(kx+\omega t)} \quad , \quad \omega = \frac{E}{\hbar}$$

The first term is a traveling wave moving in the $+x$ direction with phase velocity ω/k and the second term is a traveling wave moving in the $-x$ direction with a phase velocity ω/k.

We can think of this solution in region I as representing an incident wave moving in the $+x$ direction with amplitude A and a reflected wave moving in the $-x$ direction with an amplitude B.

The reflection takes place at the potential discontinuity much the same as a reflection occurs for light on the interface between air and glass.

We will show this clearly using wave packets later in this section.

The wave packets will also allow us to relate these results to particle motion.

In region II, $V(x) = V_0$ and Schrodinger's equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{II}(x)}{dx^2} = (E - V_0)\psi_{II}(x)$$

If we define

$$\gamma^2 = \frac{2m}{\hbar^2}(E - V_0)$$

We then again have two possible solutions

$$\psi_{II}(x) = e^{\pm i\gamma x}$$

Two cases arise

$$E > V_0 \rightarrow \gamma \text{ is real} \rightarrow \text{traveling wave solutions}$$

$$E < V_0 \rightarrow \gamma \text{ is imaginary} \rightarrow \text{real exponential solutions}$$

We first consider $E > V_0$, γ real. We then have the general solution in region II

$$\psi_{II}(x) = Ce^{+i\gamma x} + De^{-i\gamma x}$$

Now the actual physical experiment we are considering must always restrict the possibilities (always remember that we are physicists and not just solving an abstract mathematical problem).

If we assume the incident wave is traveling in the $+x$ direction in region I (coming in from $x = -\infty$), then the existence of a wave in region I traveling in the $-x$ direction makes sense.

It is a reflected wave (reflection occurs at $x = 0$).

In region II, however, we cannot physically have a wave traveling in the $-x$ direction (what is its source?).

We can, however, have a wave traveling in the $+x$ direction.

It is the transmitted wave (again think of an analogy to an air-glass interface).

So, on physical grounds, we will assume that $D = 0$.

We then have the solutions for $E > V_0$, γ real

$$\psi_I(x) = Ae^{+ikx} + Be^{-ikx} \qquad x < 0 \quad , \quad k^2 = \frac{2m}{\hbar^2}E$$

$$\psi_{II}(x) = Ce^{+i\gamma x} \qquad x > 0 \quad , \quad \gamma^2 = \frac{2m}{\hbar^2}(E - V_0)$$

We now use the continuity conditions at $x = 0$.

$$\psi_I(0) = \psi_{II}(0) \text{ and } \frac{d\psi_I(0)}{dx} = \frac{d\psi_{II}(0)}{dx}$$

which give the equations

$$A + B = C \text{ and } ik(A - B) = i\gamma C$$

which imply the solutions

$$\frac{C}{A} = \frac{2k}{k + \gamma} \text{ and } \frac{B}{A} = \frac{k - \gamma}{k + \gamma}$$

We can see the physical content by looking more deeply at the probability ideas associated with the Schrodinger equation.

For a one-particle state, the Schrodinger equation formalism says that

$$\int_\Omega \psi^*(x)\psi(x)\, d^3x = \text{ probability particle is located in volume } \Omega$$

The time rate of change of this probability is

$$\frac{\partial}{\partial t}\int_\Omega \psi^*(x)\psi(x)\,d^3x = \int_\Omega \left[\psi^*\frac{\partial\psi}{\partial t} + \psi\frac{\partial\psi^*}{\partial t}\right] d^3x$$
$$= \frac{i\hbar}{2m}\int_\Omega \left[\psi^*\nabla^2\psi - \psi\nabla^2\psi^*\right] d^3x$$
$$= \frac{i\hbar}{2m}\int_\Omega \nabla\cdot\left[\psi^*\nabla\psi - \psi\nabla\psi^*\right] d^3x$$

where we have used the time-dependent Schrodinger equation and its complex conjugate

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V(\vec{x})\psi = i\hbar\frac{\partial\psi}{\partial t} \text{ and } -\frac{\hbar^2}{2m}\nabla^2\psi^* + V(\vec{x})\psi^* = -i\hbar\frac{\partial\psi^*}{\partial t}$$

Since the volume Ω is arbitrary, this result implies a continuity equation of the form

$$\frac{\partial}{\partial t}|\psi(\vec{x},t)|^2 + \nabla\cdot\vec{J}(\vec{x},t) = 0$$

where

$$\vec{J}(\vec{x},t) = -\frac{i\hbar}{2m}\left[\psi^*\nabla\psi - \psi\nabla\psi^*\right]$$

is called the *probability flux vector* or the *probability current density*.

Using the barrier solutions for $E > V_0$, γ real, we have (for x-components)

$$J_I(x,t) = \frac{i\hbar}{2m}(A^*e^{-ikx} + B^*e^{ikx})(ikAe^{ikx} - ikBe^{-ikx})$$
$$- \frac{i\hbar}{2m}(Ae^{ikx} + Be^{-ikx})(ikA^*e^{-ikx} - ikB^*e^{ikx})$$
$$= \frac{\hbar k}{m}\left[|A|^2 - |B|^2\right] = J_{I+} - J_{I-}$$

$$J_{II}(x,t) = \frac{i\hbar}{2m}\left[(C^* e^{-i\gamma x})(i\gamma C e^{i\gamma x}) - (C e^{i\gamma x})(-i\gamma C^* e^{-i\gamma x})\right]$$
$$= \frac{\hbar\gamma}{m}|C|^2 = J_{II+}$$

We then define the reflection and transmission probabilities in terms of the current as

$$R = \frac{J_{I-}}{J_{I+}} = \frac{|B|^2}{|A|^2} \text{ and } T = \frac{J_{II+}}{J_{1+}} = \frac{\gamma}{k}\frac{|C|^2}{|A|^2}$$

Now, physically, we identify

$$|A|^2 = \text{incident beam intensity}$$
$$|B|^2 = \text{reflected beam intensity}$$
$$|C|^2 = \text{transmitted beam intensity}$$

and therefore

R = probability that an incident wave will be reflected

T = probability that an incident wave will be transmitted

We find

$$R = \frac{(k-\gamma)^2}{(k+\gamma)^2} \text{ and } T = \frac{4k\gamma}{(k+\gamma)^2}$$

Notice that

$$R + T = 1$$

which makes physical sense since the total probability for the wave to go some- where must equal one.

In the book, I now recast the case $E > V_0$ in terms of the wave packet formalism derived earlier.

This allows us to make particle interpretations.

I will skip that section in these lectures.

Now what happens if $E = p^2/2m < V_0$?

We then have the solutions (let $A = 1$)

$$x < 0 \qquad \psi_I = e^{ikx} + Re^{-ikx}$$

$$x > 0 \qquad \psi_{II} = Se^{-\beta x}$$

where $\hbar k = \sqrt{2mE}$ and $\hbar\beta = \sqrt{2m(V_0 - E)}$.

We have excluded the mathematical solution $e^{\beta x}$ for $x > 0$ because it would cause the integral

$$\int_{-\infty}^{\infty} |\psi|^2 \, dx$$

to diverge, which means it is a *non-normalizable* or *unphysical solution*.

This is an example of a solution not being well-behaved.

At $x = 0$, we have the continuity conditions which give

$$1 + R = S \text{ and } 1 - R = -\frac{\beta}{ik}S \text{ or}$$

$$S = \frac{2ik}{ik - \beta} \text{ and } R = \frac{ik + \beta}{ik - \beta}$$

Since the solution for $x > 0$ is not a traveling wave, there is no transmitted packet.

The incident wave packet analysis is the same as before.

The reflected wave packet analysis is different because $R(p)$ is not a real function.

In fact,

$$|R|^2 = 1 \rightarrow R = e^{i\phi} = \frac{ip + \sqrt{2mV_0 - p^2}}{ip - \sqrt{2mV_0 - p^2}}$$

$$= \frac{ip + \sqrt{2mV_0 - p^2}}{ip - \sqrt{2mV_0 - p^2}} \frac{ip + \sqrt{2mV_0 - p^2}}{ip + \sqrt{2mV_0 - p^2}}$$

or

$$e^{i\phi} = \frac{p^2 - mV_0 - ip\sqrt{2mV_0 - p^2}}{mV_0} = \cos\phi + i\sin\phi$$

with

$$\cos\phi = \frac{p^2 - mV_0}{mV_0} \text{ and } \sin\phi = -\frac{p\sqrt{2mV_0 - p^2}}{mV_0}$$

Therefore, we have as wave packet representations

$$\psi_{inc}(x,t) = \int_0^\infty \frac{dp}{2\pi\hbar} f(p) f(p) e^{i(px - Et)/\hbar} \text{ where } E = \frac{p^2}{2m},\ p = \hbar k$$

$$\psi_{refl}(x,t) = \int_0^\infty \frac{dp}{2\pi\hbar} f(p) f(p) e^{-i(px + Et - \hbar\phi(p))/\hbar}$$

$$\text{where } E = \frac{p^2}{2m},\ p = \hbar k$$

The stationary phase argument implies

$$x = \frac{p_0}{m} t = v_0 t \text{ for the incident packet}$$

which says that the incident packet arrives at x = 0 at t = 0 and

$$x = -\frac{p_0}{m}t + \hbar \left.\frac{\partial\phi(p)}{\partial p}\right|_{p=p_0} \quad \text{for the reflected packet}$$

This says that the reflected wave packet leaves $x = 0$ at

$$t = t_{delay} = \frac{\hbar m}{p_0} \left.\frac{\partial\phi(p)}{\partial p}\right|_{p=p_0}$$

which is NOT the same time that the incident packet arrived!

Continuing we have

$$\frac{\partial \cos\phi(p)}{\partial p} = -\sin\phi(p)\frac{\partial\phi(p)}{\partial p} = \frac{2p}{mV_0}$$

$$\frac{\partial\phi(p)}{\partial p} = \frac{2}{\sqrt{2mV_0 - p_0^2}}$$

which gives

$$t_{delay} = \frac{m\hbar}{p_0}\frac{2}{\sqrt{2mV_0 - p_0^2}} = \frac{2m}{p_0}\frac{1}{\beta}$$

Now, we found that

$$\psi_{II}(x) = Se^{-\beta x}$$

which says that the probability amplitude is significantly different from zero up to a distance $d = 1/\beta$ approximately.

In other words,

$$\psi_{II}(x = 1/\beta) = \frac{1}{e}\psi_{II}(x = 0) \rightarrow |\psi_{II}(x = 1/\beta)|^2 \approx \frac{1}{7}|\psi_{II}(x = 0)|^2$$

Therefore, to a good approximation, we have the striking result $t_{\text{delay}} = 2d/v_0$.

What does this mean?

It seems as if there is a nonzero probability for finding the particle in the classically forbidden region $x > 0$, where K = kinetic energy = $E - V_0 < 0$ since $\psi_{\text{II}}(x) = e^{-\beta x} \neq 0$ there.

Since K cannot be less than zero, it seems like we must have a violation of conservation of energy if the particle were actually found in the region $x > 0$.

This is NOT the case, however.

If we observe the particle in this forbidden region, then it will no longer be in a state with $E < V_0$.

The act of measuring the location of the particle must necessarily introduce an uncertainty in p and hence in E.

The particle seems to have an appreciable probability to exist in the region up to a distance $d = 1/\beta$.

If we observe the particle on the right, we have then localized it such that $\Delta x \approx 1/\beta$.

This says that we have introduced an uncertainty in p

$$\Delta p \geq \frac{\hbar}{\Delta x} \approx \hbar\beta$$

and a corresponding energy uncertainty

$$\Delta E \approx \frac{(\Delta p)^2}{2m} \geq \frac{\hbar^2\beta^2}{2m} = V_0 - E$$

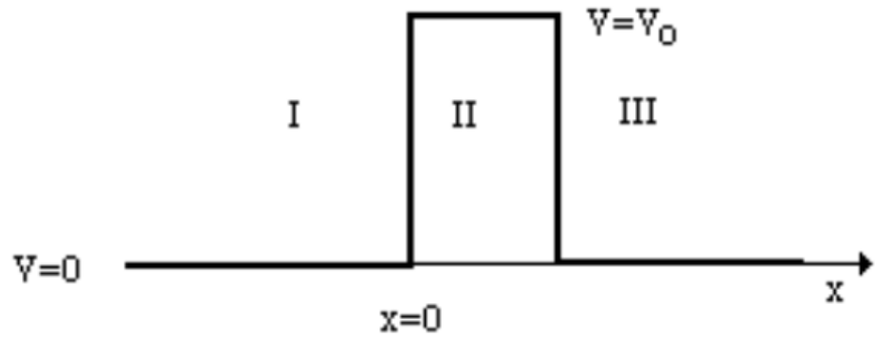
This implies that E is now uncertain enough that we can no longer claim that energy conservation has been violated!

Quantum mechanics has a way of covering its own tracks!

Tunneling

We now change the potential energy function so that the barrier (step) is not infinitely thick (which prevented a traveling wave from existing for $E < V_0$, $x > 0$.

The new potential energy function is shown in the Figure below.



Finite Step - Finite Width barrier

For $E > V_0$ the results are similar to infinitely thick barrier and no new physical ideas appear.

For $E < V_0$, however, we get some very interesting new physical results.

As we shall see, it turns out that a real traveling wave can appear on the other side of the barrier (even though there are no sources on that side) in this case.

This is called quantum tunneling.

Let us see how it works.

We have three regions I, II and III to consider as shown in the Figure.

We get the solutions

$$x \le 0 \qquad -\frac{\hbar^2}{2m}\frac{d^2\psi_I}{dx^2} = E\psi_I$$

$$\psi_I = A_1 e^{ikx} + B_1 e^{-ikx} \quad , \quad E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \ , \ k \text{ real}$$

$$0 \le x \le a \qquad -\frac{\hbar^2}{2m}\frac{d^2\psi_{II}}{dx^2} + V_0\psi_{II} = E\psi_{II}$$

$$\psi_{II} = Ce^{\gamma x} + De^{-\gamma x} \quad , \quad V_0 - E = \frac{p^2}{2m} = \frac{\hbar^2\gamma^2}{2m} \ , \ \gamma \text{ real}$$

$$x \ge a \qquad -\frac{\hbar^2}{2m}\frac{d^2\psi_{III}}{dx^2} = E\psi_{III}$$

$$\psi_{III} = A_2e^{ikx} + B_2e^{-ikx} \quad , \quad E = \frac{p^2}{2m} = \frac{\hbar^2k^2}{2m} \ , \ k \text{ real}$$

The probability current does not vanish at $x = \pm\infty$, which implies that we must assume the existence of distant sources and sinks of probability(or particles).

We have two sets of continuity equations (at $x = 0$ and $x = a$).

At $x = 0$ we get

$$\psi_I(0) = \psi_{II}(0) \rightarrow A_1 + B_1 = C + D$$

$$\frac{d\psi_I(0)}{dx}\frac{d\psi_{II}(0)}{dx} \rightarrow ik(A_1 - B_1) = \gamma(C - D)$$

and at $x = a$ we get

$$\psi_{II}(a) = \psi_{III}(a) \rightarrow Ce^{\gamma a} + De^{-\gamma a} = A_2e^{ika} + B_2e^{-ika}$$

$$\frac{d\psi_{II}(a)}{dx}\frac{d\psi_{III}(a)}{dx} \rightarrow \gamma(Ce^{\gamma a} - De^{-\gamma a}) = ik(A_2e^{ika} - B_2e^{-ika})$$

We can restate these equations in matrix form.

At $x = 0$ we get

$$\begin{pmatrix} 1 & 1 \\ ik & -ik \end{pmatrix}\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \hat{M}_1 \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \gamma & -\gamma \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix} = \hat{M}_2 \begin{pmatrix} C \\ D \end{pmatrix}$$

and at $x = a$ we get

$$\begin{pmatrix} e^{\gamma a} & e^{-\gamma a} \\ \gamma e^{\gamma a} & -\gamma e^{-\gamma a} \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix} = \hat{M}_3 \begin{pmatrix} C \\ D \end{pmatrix}$$

$$= \begin{pmatrix} e^{ika} & e^{-ika} \\ ike^{ika} & -ike^{-ika} \end{pmatrix}\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} = \hat{M}_4 \begin{pmatrix} A_2 \\ B_2 \end{pmatrix}$$

The transmission/reflection properties of the barrier are given by the coefficients A_1, B_1, A_2, B_2, which we assume are related by a transfer matrix $\hat{Y}$, where

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \hat{Y}\begin{pmatrix} A_2 \\ B_2 \end{pmatrix}$$

We then have

$$\hat{M}_1 \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \hat{M}_2 \begin{pmatrix} C \\ D \end{pmatrix} \rightarrow \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \hat{M}_1^{-1}\hat{M}_2 \begin{pmatrix} C \\ D \end{pmatrix}$$

$$\hat{M}_3 \begin{pmatrix} C \\ D \end{pmatrix} = \hat{M}_4 \begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \rightarrow \begin{pmatrix} C \\ D \end{pmatrix} = \hat{M}_3^{-1}\hat{M}_4 \begin{pmatrix} A_2 \\ B_2 \end{pmatrix}$$

or

$$\hat{Y} = \hat{M}_1^{-1}\hat{M}_2\hat{M}_3^{-1}\hat{M}_4$$

A great deal of algebra gives

$$Y_{11} = e^{ika}\left[\cosh\gamma a + \frac{1}{2}i\sinh\gamma a\left(\frac{\gamma}{k} - \frac{k}{\gamma}\right)\right]$$

$$Y_{12} = \frac{1}{2}e^{-ika}\sinh\gamma a\left(\frac{\gamma}{k} + \frac{k}{\gamma}\right)$$

$$Y_{21} = -\frac{1}{2}ie^{ika}\sinh\gamma a\left(\frac{\gamma}{k} + \frac{k}{\gamma}\right)$$

$$Y_{22} = e^{-ika}\left[\cosh\gamma a - \frac{1}{2}i\sinh\gamma a\left(\frac{\gamma}{k} - \frac{k}{\gamma}\right)\right]$$

If we have sources only on the left of the barrier, then we must choose $B_2 = 0$.

The currents are

$$J_{LEFT}(x,t) = \frac{\hbar k}{m}\left[|A_1|^2 - |B_1|^2\right]$$
$$= J_{LEFT+}(x,t) - J_{LEFT-}(x,t) \qquad x < 0$$

$$J_{RIGHT}(x,t) = J_{RIGHT+}(x,t) = \frac{\hbar k}{m}|A_2|^2 \qquad x > 0$$

The reflection and transmission probabilities are given by

$$R = \frac{|B_1|^2}{|A_1|^2} = \frac{|Y_{21}|^2}{|Y_{11}|^2} \quad , \quad T = \frac{|A_2|^2}{|A_1|^2} = \frac{1}{|Y_{11}|^2}$$

Algebra shows $R+T = 1$ as it must in order to conserve probability (or particles).

Evaluating the expression for T we get

$$T = \frac{1}{1 + \frac{V_0^2\sinh^2\gamma a}{4E(V_0 - e)}} \quad , \quad \gamma^2 = \frac{2m}{\hbar^2}(V_0 - E)$$

The fact that $T > 0$ for $E < V_0$ implies the existence of tunneling.

The probability amplitude leaks through the barrier.

It is important to realize that the fact that $T > 0$, **DOES NOT** say that particles or wave packets passed through the barrier.

No measurement can be done on the system that will allow us to observe a particle in the region $0 < x < a$ with $E < V_0$, since this would violate energy conservation.

It is **ONLY** probability that is leaking through.

If this causes the probability amplitude and hence the probability to be nonzero on the other side of the barrier, than it *must be possible for us to observe* the particle on the other side, i.e., we can observe the particle on the left side of the barrier with $E < V_0$, but we can never observe it in the region of the barrier with $E < V_0$.

That is what is being said here.

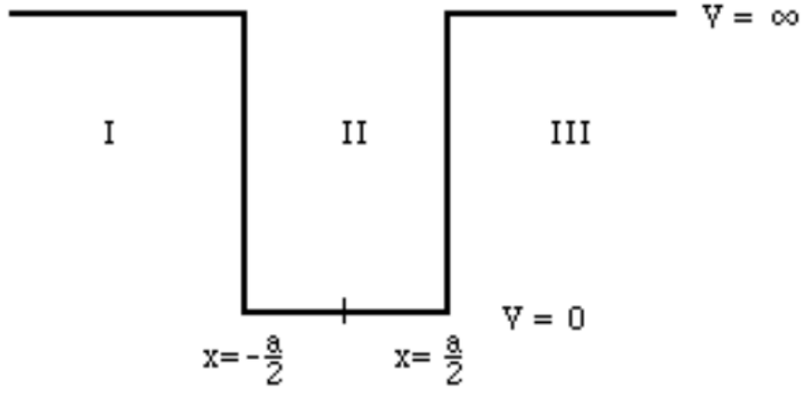
Bound States

Infinite Square Well

We now consider the potential energy function

$$V(x) = \begin{cases} 0 & -\frac{a}{2} \le x \le \frac{a}{2} \\ \infty & |x| \ge \frac{a}{2} \end{cases}$$

This is the so-called *infinite square well* shown in the Figure.



Infinite Square Well

This is an example of a potential that is infinite in an extended region.

Therefore, we must require that the wave function $\psi(x) = 0$ in these regions or the Schrodinger equation makes no sense mathematically.

In this case we have

$$\psi_I(x) = 0 \text{ and } \psi_{III}(x) = 0$$

In region II we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{II}}{dx^2} = E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}$$

and

$$\psi_{II}(x) = Ae^{ikx} + Be^{-ikx}$$

The continuity of the wavefunction at $x = \pm a/2$ says that we must have

$$\psi_{II}\left(-\frac{a}{2}\right) = Ae^{-ika/2} + Be^{ika/2} = 0$$

$$\psi_{II}\left(\frac{a}{2}\right) = Ae^{ika/2} + Be^{-ika/2} = 0$$

which imply that

$$\frac{B}{A} = -e^{-ika} = -e^{ika}$$

This is an equation for the allowed values(values corresponding to a valid solution) of k.

This equation is

$$e^{2ika} = 1$$

The allowed values of *k* form a discrete spectrum of energy eigenvalues (quantized energies) given by

$$2k_n a = 2n\pi \rightarrow k_n = \frac{n\pi}{a} \rightarrow E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{n^2\pi^2\hbar^2}{2ma^2}\ ,\ n = 1, 2, 3, 4, \ldots$$

The corresponding wave functions are

$$\begin{aligned}\psi_{II}^{(n)}(x) &= A_n(e^{ik_n x} - e^{-ik_n a}e^{-ik_n x}) \\ &= A_n e^{-ik_n a/2}(e^{ik_n(x+a/2)} - e^{-ik_n(x+a/2)}) \\ &= \tilde{A}_n \sin k_n(x + a/2)\end{aligned}$$

where $\tilde{A}_n$ is determined bv the normalization condition

$$\int_{-a/2}^{a/2} |\psi_n(x)|^2\, dx = 1$$

Substituting the value of k_n we get

$$\begin{aligned}\psi_{II}^{(1)}(x) &= \tilde{A}_1 \sin k_1(x + a/2) = \tilde{A}_1 \sin \pi(x + a/2)/a \\ &= \tilde{A}_1 \sin(\pi x/a + \pi/2) = \tilde{A}_1 \cos(\pi x/a)\end{aligned}$$

$$\begin{aligned}\psi_{II}^{(2)}(x) &= \tilde{A}_2 \sin k_2(x + a/2) = \tilde{A}_2 \sin 2\pi(x + a/2)/a \\ &= \tilde{A}_2 \sin(2\pi x/a + \pi) = \tilde{A}_2 \sin(2\pi x/a)\end{aligned}$$

$$\begin{aligned}\psi_{II}^{(3)}(x) &= \tilde{A}_3 \sin k_3(x + a/2) = \tilde{A}_3 \sin 3\pi(x + a/2)/a \\ &= \tilde{A}_3 \sin(3\pi x/a + 3\pi/2) = \tilde{A}_3 \cos(3\pi x/a)\end{aligned}$$

or

$$\psi_{II}(x) = \begin{cases} \sin(n\pi x/a) & \text{n even} \\ \cos(n\pi x/a) & \text{n odd} \end{cases} \qquad n = 1, 2, 3, 4, \ldots$$

We have mathematically solved the ordinary differential equation problem, now what is the physical meaning of these results?

We find a discrete spectrum of allowed energies corresponding to bound states of the Hamiltonian; *the energy is quantized*.

Bound states designate states which are localized in space, i.e., the probability is large only over restricted regions of space and goes to zero far from the potential region.

The *lowest* energy value or lowest energy *level* or *ground state energy* is

$$E_1 = \frac{\pi^2 \hbar^2}{2ma^2} > 0$$

with

$$\psi_1(x) = \begin{cases} A\cos(n\pi x/a) & |x| \le \frac{a}{2} \\ 0 & |x| \ge \frac{a}{2} \end{cases} \qquad n = 1, 2, 3, 4, \ldots$$

This minimum energy is not zero because of the Heisenberg uncertainty principle.

Since the particle has a nonzero amplitude for being in the well, we say that it is localized such that $\Delta x \approx a$ and thus

$$\Delta p \ge \frac{\hbar}{\Delta x} \approx \frac{\hbar}{a}$$

This says that the kinetic energy (or energy in this case because the potential energy equals zero in region II) must have a minimum value given approximately by

$$E_{min} = K_{min} \approx \frac{(\Delta p)^2}{2m} \approx \frac{\hbar^2}{2ma^2}$$

Let us look more closely at the wave functions.

The integer $n-1$ corresponds to the number of nodes (zeros) of the wave function (other than the well edges).

They also have the property

$$\psi(-x) = \psi(x) \text{ for n odd and } \psi(-x) = -\psi(x) \text{ for n even}$$

The above discrete transformation of the wave function corresponds to the *parity* operator $\hat{\mathcal{P}}$ where we have

$$\hat{\mathcal{P}}\psi(x) = \psi(-x) = \psi(x) \text{ means even parity}$$

$$\hat{\mathcal{P}}\psi(x) = \psi(-x) = -\psi(x) \text{ means odd parity}$$

Let us look more generally at the parity operation.

Suppose that the potential energy function obeys the rule $V(\vec{x}) = V(-\vec{x})$ and let $\psi(\vec{x})$ be a solution of the Schrodinger equation with energy E

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{x})\right)\psi(\vec{x}) = E\psi(\vec{x})$$

Now let $\vec{x} \rightarrow -\vec{x}$ to get the equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(-\vec{x})\right)\psi(-\vec{x}) = E\psi(-\vec{x})$$

or

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{x})\right)\psi(-\vec{x}) = E\psi(-\vec{x})$$

This says that, if $\psi(\vec{x})$ is a solution of the Schrodinger equation with energy E, then $\psi(-\vec{x})$ is also a solution of the Schrodinger equation with the same energy E.

This says that the combinations

$$\psi(\vec{x}) \pm \psi(-\vec{x})$$

are also solutions of the Schrodinger equation with the same energy E.

Now

$$\psi(\vec{x}) + \psi(-\vec{x}) \rightarrow \text{ an even parity solution}$$
$$\psi(\vec{x}) - \psi(-\vec{x}) \rightarrow \text{ an odd parity solution}$$

This says that if $V(\vec{x}) = V(-\vec{x})$ (a symmetric potential), then we can always choose solutions that have a definite parity (even or odd).

We formally define the *parity operator* by the relation

$$\langle \vec{x} | \hat{\mathcal{P}} | \psi \rangle = \langle -\vec{x} | \psi \rangle$$

Since

$$\langle \vec{x} | \hat{\mathcal{P}}^2 | \psi \rangle = \langle -\vec{x} | \hat{\mathcal{P}} | \psi \rangle = \langle \vec{x} | \psi \rangle$$

we must have

$$\hat{\mathcal{P}}^2 = \hat{I}$$

which means the eigenvalues of are ±1 as we indicated earlier.

This also says that $\hat{\mathcal{P}}^{-1} = \hat{\mathcal{P}}$.

We can show $[\ \hat{H}\ ,\ \hat{\mathcal{P}}\ ] = 0$ for symmetric potentials by

$$\hat{\mathcal{P}}\hat{H}\,|E\rangle = \hat{\mathcal{P}}E\,|E\rangle = E\hat{\mathcal{P}}\,|E\rangle = \pm E\,|E\rangle$$

$$\hat{H}\hat{\mathcal{P}}\,|E\rangle = \pm\hat{H}\,|E\rangle = \pm E\,|E\rangle$$

$$(\hat{\mathcal{P}}\hat{H} - \hat{H}\hat{\mathcal{P}})\,|E\rangle = 0$$

$$[\hat{H}, \hat{\mathcal{P}}] = 0$$

since $|E\rangle$ is an arbitrary state.

As we saw earlier, this commutator relationship says that

$$\hat{H}\hat{\mathcal{P}} = \hat{\mathcal{P}}\hat{H}$$

$$\hat{\mathcal{P}}\hat{H}\hat{\mathcal{P}} = \hat{\mathcal{P}}^2\hat{H} = \hat{H}$$

$$\hat{\mathcal{P}}^{-1}\hat{H}\hat{\mathcal{P}} = \hat{H}$$

which means that $\hat{H}$ is invariant under the $\hat{\mathcal{P}}$ transformation.

We have used $\hat{\mathcal{P}}^2 = \hat{I}$ in this derivation.

It also says that

$$\hat{H}(\hat{\mathcal{P}}\,|E\rangle) = \hat{\mathcal{P}}\hat{H}\,|E\rangle = E(\hat{\mathcal{P}}\,|E\rangle)$$

or $\hat{\mathcal{P}}\,|E\rangle$ is an eigenstate of $\hat{H}$ with energy E as we stated above.

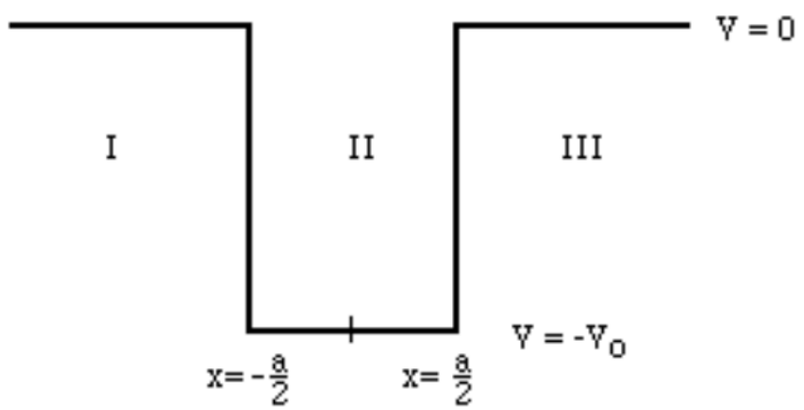
The concept of parity invariance and the fact that $\hat{H}$ and $\hat{\mathcal{P}}$ share a common set of eigenfunctions can greatly simplify the solution of the Schrodinger equation in many cases.

The Finite Square Well

We now consider the potential energy function

$$V(x) = \begin{cases} -V_0 & |x| \le \frac{a}{2} \\ 0 & |x| \ge \frac{a}{2} \end{cases}$$

This is the so-called *finite square well* (in one dimension) and it is shown in the Figure below.



Finite Square Well

The solutions are:

Region I : $x < \frac{a}{2}$

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_I}{dx^2} = E\psi_I \ , 0 \ge E \ge -V_0 \ , \hbar^2 k^2 = 2m|E| \ , E = -|E|$$

with solutions

$$\psi_I(x) = Ae^{-kx} + Be^{kx}$$

Since $x = -\infty$ is included in this region, we must exclude the e^{-kx} term by choosing $A = 0$, which gives

$$\psi_I(x) = Be^{kx} \qquad x < -\frac{a}{2}$$

Region II: $-\frac{a}{2} \leq x \leq \frac{a}{2}$

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{II}}{dx^2} - V_0\psi_{II} = E\psi_{II}\ , 0 \geq E \geq -V_0\ , p^2 = 2m(V_0 - |E|)\ , E = -|E|$$

with solutions

$$\psi_{II}(x) = Ce^{ipx/\hbar} + De^{-ipx/\hbar}$$

Region III: $x > \frac{a}{2}$

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{III}}{dx^2} = E\psi_{III}\ , 0 \geq E \geq -V_0\ , \hbar^2k^2 = 2m|E|\ , E = -|E|$$

with solutions

$$\psi_{III}(x) = Fe^{kx} + Ge^{-kx}$$

Since x = +∞ is included in this region, we must exclude the e^{kx} term by choosing $F = 0$, which gives

$$\psi_{III}(x) = Ge^{-kx}$$

These results represent a general solution to the problem.

There seems to be 4 unknown constants, namely, B, C, D, and G.

However, since $V(x) = V(-x)$, parity is conserved and we can choose even and odd solutions, or solutions of definite parity.

Even parity implies $\psi(x) = \psi(-x)$ or $G = B$ and $C = D$.

This solution is

$$\psi(x) = \begin{cases} C\cos px/\hbar & |x| \le \frac{a}{2} \\ Be^{-kx} & x > \frac{a}{2} \\ Be^{kx} & x < -\frac{a}{2} \end{cases}$$

Odd parity implies $\psi(x) = -\psi(-x)$ or $G = -B$ and $D = -C$.

This solution is

$$\psi(x) = \begin{cases} C\sin px/\hbar & |x| \le \frac{a}{2} \\ Be^{-kx} & x > \frac{a}{2} \\ -Be^{kx} & x < -\frac{a}{2} \end{cases}$$

Thus, by using parity we reduce the number of unknowns in the problem to two for each type of solution.

We now impose the continuity conditions of the wave function and its derivative only at $x = a/2$ for both solutions.

Since these are definite parity solutions the continuity condition at $x = -a/2$ will give no new information and is not needed.

Even Parity Results

$$C\cos pa/2\hbar = Be^{-ka/2} \text{ and } -\frac{p}{\hbar}C\sin pa/2\hbar = -kBe^{-ka/2}$$

or

$$\frac{B}{C} = e^{ka/2}\cos pa/2\hbar = \frac{p}{\hbar k}e^{ka/2}\sin pa/2\hbar$$

so that

$$p\tan pa/2\hbar = \hbar k$$

This last equation is a transcendental equation for E and its solutions determine the allowed E values for the even parity states for this potential energy function.

These E values are the even parity energies or energy levels of a particle in the finite square well potential.

Odd Parity Results

$$C\sin pa/2\hbar = Be^{-ka/2} \text{ and } \frac{p}{\hbar}C\cos pa/2\hbar = -kBe^{-ka/2}$$

or

$$\frac{B}{C} = e^{ka/2}\sin pa/2\hbar = \frac{p}{\hbar k}e^{ka/2}\cos pa/2\hbar$$

so that

$$p\cot pa/2\hbar = -\hbar k$$

Again, this last equation is a transcendental equation for E and its solutions determine the allowed E values for the odd parity states for this potential energy function.

These E values are the odd parity energies or energy levels of a particle in the finite square well potential.

In general, at this stage of the solution, we must either devise a clever numerical or graphical trick to find the solutions of the transcendental equations or resort to a computer.

The first thing one should always do is change variables to get rid of as many extraneous constants as possible.

In this case we let

$$\beta = ka = \frac{a}{\hbar}\sqrt{2m|E|} \quad , \quad \alpha = \gamma a = \frac{p}{\hbar}a = \frac{a}{\hbar}\sqrt{2m(V_0 - |E|)}$$

The first useful equation we can derive is

$$\alpha^2 + \beta^2 = \frac{2mV_0a^2}{\hbar^2} = \text{ constant for a given well}$$

This is the equation of a circle of radius

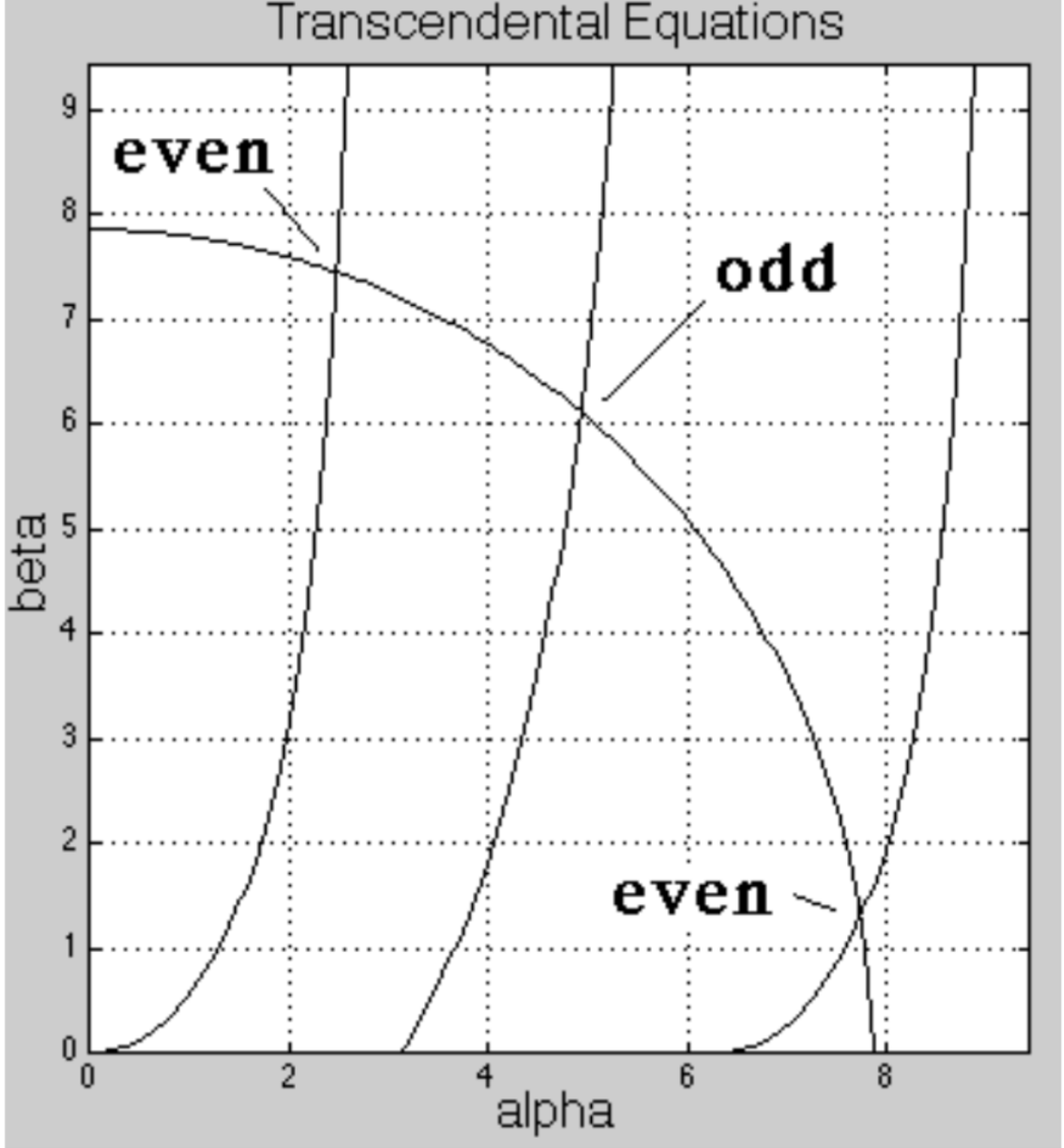
$$\sqrt{\frac{2mV_0a^2}{\hbar^2}}$$

With these new variables the two transcendental equations are

$$\beta = \alpha \tan\frac{\alpha}{2} \text{ for even parity and } \beta = -\alpha \cot\frac{\alpha}{2} \text{ for odd parity}$$

We can find solutions graphically by plotting as shown in the Figure on the next slide for the case (effectively a choice of the quantity $V_0\, a^2$)

$$\text{circle radius } = \sqrt{\frac{2mV_0a^2}{\hbar^2}} = \frac{5\pi}{2}$$



Solutions for $\sqrt{\frac{2mV_0a^2}{\hbar^2}} = \frac{5\pi}{2}$

The solutions correspond to the intersections of the circle (fixed for a given well) and the curves represented by the two transcendental equations.

They are shown in the Figure.

For the choice of potential well shown in the figure we have 2 even parity solutions and 1 odd parity solution.

These correspond to the allowed energy levels for this particular well and the corresponding wave functions and energies represent bound states of the well.

We can also do a straight numerical solution for even parity by rearranging the equations as follows:

$$\alpha^2 + \beta^2 = \frac{2mV_0a^2}{\hbar^2} \text{ and } \beta = \alpha \tan\frac{\alpha}{2}$$

$$\alpha^2\left(1 + \tan^2\frac{\alpha}{2}\right) = \frac{2mV_0a^2}{\hbar^2}$$

$$\alpha^2 - \frac{25\pi^2}{4}\cos^2\frac{\alpha}{2} = f(\alpha) = 0$$

The numerical solution of this equation can be carried out by any standard technique (Newton-Raphson method, for instance) for finding the zeros of the function f(α).

For this case we get

$$\alpha = 2.4950 \text{ and } 7.1416$$

which is clearly in agreement with the graphical result.

Transmission Resonances

As we have seen the spectrum of energies for the square well is made up of a finite number of bound-state levels.

The most important feature of a bound-state wave function is that it is localized in space, i.e., it falls off exponentially as we move away from the well.

Another interesting feature of the square well has to do with the continuous part of its energy spectrum.

All energy states with $E > 0$ are allowed.

Again there are three regions to consider, labeled in the same way as earlier.

We will again assume that we have a wave incident from the left (from $x = -\infty$) with unit intensity

$$\psi_I(x) = e^{ikx} + Be^{-ikx} \text{ where } E = \frac{\hbar^2 k^2}{2m}$$

where B is the amplitude for the reflected wave.

In region II, which is over the well, we have the solution

$$\psi_{II}(x) = Ce^{ik_1 x} + De^{-ik_1 x} \text{ where } E + V_0 = \frac{\hbar^2 k_1^2}{2m}$$

In region III, we only have a transmitted wave as the solution

$$\psi_{III}(x) = Fe^{ikx} \text{ where } E = \frac{\hbar^2 k^2}{2m}$$

Again, we must match the wave functions and their derivatives at $x = \pm a/2$.

This gives us 4 equations for the 4 unknown coefficients.

Usually an author now says that *lots of algebra* gives us the transmission coefficient and writes down the answer.

Let us actually do it just once.

Continuity of the wave function and its derivative at $x = -a/2$ gives

$$e^{-ika/2} + Be^{ika/2} = Ce^{-ik_1 a/2} + De^{ik_1 a/2}$$
$$ike^{-ika/2} - ikBe^{ika/2} = ik_1 Ce^{-ik_1 a/2} - ik_1 De^{ik_1 a/2}$$

Continuity of the wave function and its derivative at $x = +a/2$ gives

$$Fe^{ika/2} = Ce^{ik_1 a/2} + De^{-ik_1 a/2}$$
$$ikFe^{ika/2} = ik_1 Ce^{ik_1 a/2} - ik_1 De^{-ik_1 a/2}$$

Then

$$ik_1 e^{-ika/2} + ik_1 Be^{ika/2} = ik_1 Ce^{-ik_1 a/2} + ik_1 De^{ik_1 a/2}$$
$$ike^{-ika/2} - ikBe^{ika/2} = ik_1 Ce^{-ik_1 a/2} - ik_1 De^{ik_1 a/2}$$
$$ik_1 Fe^{ika/2} = ik_1 Ce^{ik_1 a/2} + ik_1 De^{-ik_1 a/2}$$
$$ikFe^{ika/2} = ik_1 Ce^{ik_1 a/2} - ik_1 De^{-ik_1 a/2}$$

Solving for C and D we get

$$2ik_1 Ce^{-ik_1 a/2} = i(k + k_1)e^{-ika/2} - i(k - k_1)Be^{ika/2}$$
$$2ik_1 De^{ik_1 a/2} = -i(k - k_1)e^{-ika/2} + i(k + k_1)Be^{ika/2}$$
$$2ik_1 Ce^{ik_1 a/2} = i(k + k_1)Fe^{ika/2}$$
$$2ik_1 De^{-ik_1 a/2} = -i(k - k_1)Fe^{ika/2}$$

Rearranging we have

$$C = \left(\frac{k+k_1}{2k_1}\right) Fe^{-ik_1 a/2} e^{ika/2} = e^{ik_1 a/2}\left[\left(\frac{k+k_1}{2k_1}\right) e^{-ika/2} - \left(\frac{k-k_1}{2k_1}\right) Be^{ika/2}\right]$$
$$D = -\left(\frac{k-k_1}{2k_1}\right) Fe^{ik_1 a/2} e^{ika/2} = e^{-ik_1 a/2}\left[-\left(\frac{k-k_1}{2k_1}\right) e^{-ika/2} + \left(\frac{k+k_1}{2k_1}\right) Be^{ika/2}\right]$$

Therefore

$$\left(\frac{k+k_1}{2k_1}\right) Fe^{-ik_1 a} e^{ika/2} - \left(\frac{k+k_1}{2k_1}\right) e^{-ika/2} = -\left(\frac{k-k_1}{2k_1}\right) Be^{ika/2}$$
$$-\left(\frac{k-k_1}{2k_1}\right) Fe^{ik_1 a} e^{ika/2} + \left(\frac{k-k_1}{2k_1}\right) e^{-ika/2} = \left(\frac{k+k_1}{2k_1}\right) Be^{ika/2}$$

Dividing we get

$$\frac{\left(\frac{k+k_1}{2k_1}\right)Fe^{-ik_1a}e^{ika/2}-\left(\frac{k+k_1}{2k_1}\right)e^{-ika/2}}{-\left(\frac{k-k_1}{2k_1}\right)Fe^{ik_1a}e^{ika/2}+\left(\frac{k-k_1}{2k_1}\right)e^{-ika/2}}=\frac{-\left(\frac{k-k_1}{2k_1}\right)}{\left(\frac{k+k_1}{2k_1}\right)}$$

$$(k+k_1)^2\left[Fe^{-ik_1a}e^{ika/2}-e^{-ika/2}\right]=-(k-k_1)^2\left[-Fe^{ik_1a}e^{ika/2}+e^{-ika/2}\right]$$

or

$$\left[(k+k_1)^2e^{-ik_1a}-(k-k_1)^2e^{ik_1a}\right]e^{ika/2}F$$
$$=-(k-k_1)^2e^{-ika/2}+(k+k_1)^2e^{-ika/2}$$

Finally we solve for the transmission coefficient F to get

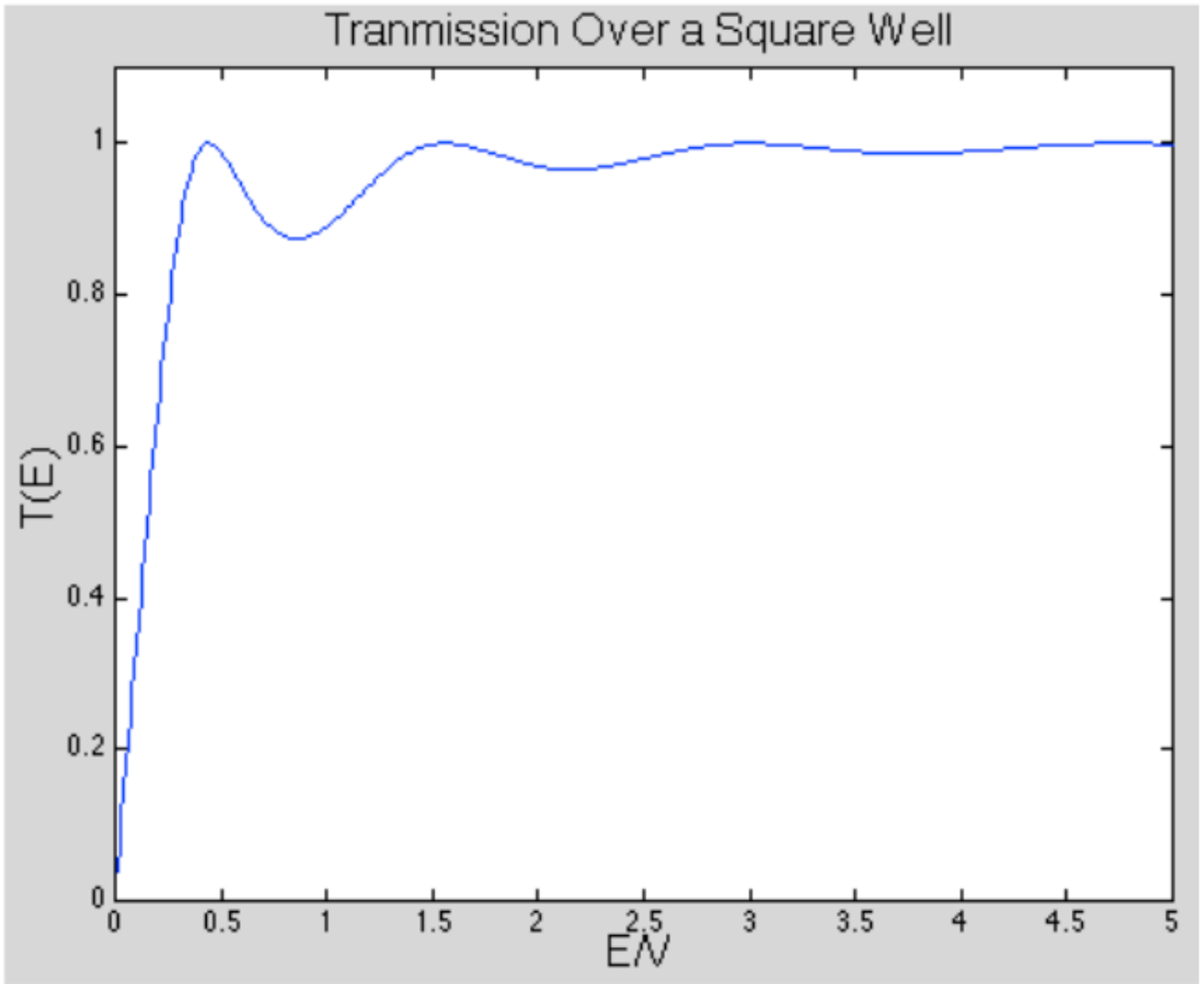
$$F=\frac{(k+k_1)^2-(k-k_1)^2}{(k+k_1)^2e^{-ik_1a}-(k-k_1)^2e^{ik_1a}}e^{-ika}$$
$$=\frac{4kk_1}{(k^2+k_1^2)(e^{-ik_1a}-e^{ik_1a})+2kk_1(e^{-ik_1a}+e^{ik_1a})}e^{-ika}$$
$$=\frac{e^{-ika}}{\cos k_1a-\frac{i}{2}\left(\frac{k}{k_1}+\frac{k_1}{k}\right)\sin k_1a}$$

and the transmission probability

$$T(E)=|F(E)|^2=\frac{1}{1+\frac{\sin^2\sqrt{\frac{2m}{\hbar^2}V_0a^2(1+\frac{E}{V_0})}}{4\left(\frac{E}{V_0}\right)\left(1+\frac{E}{V_0}\right)}}$$

A plot of $T(E)$ versus E is shown below for the case

$$\frac{2mV_0a^2}{\hbar^2} = \frac{25\pi^2}{4}$$



Transmission over a Square Well

The peaks are called *resonances*.

They occur when $T(E) = 1$ or

$$\sin^2 \sqrt{\frac{2m}{\hbar^2} V_0 a^2 (1 + \frac{E}{V_0})} = 0$$

or when

$$\sqrt{\frac{2m}{\hbar^2}V_0a^2(1+\frac{E}{V_0})} = n\pi$$
$$\frac{E}{V_0} = \frac{n^2\pi^2}{\frac{2m}{\hbar^2}V_0a^2} - 1 = \frac{4n^2}{25} - 1 \geq 0$$
$$\frac{E}{V_0} = 0.44, 1.56, 3.00, ...$$

in agreement with the diagram.

To a particle of these particular energies, the potential looks completely transparent; there is no reflected particle

$$R(E) = |B(E)|^2 = 0$$

All of the incident wave is transmitted into region III.

A special feature of these amplitudes, not derivable from non-relativistic quantum mechanics, relates the properties of the transmission probability and the bound state energies of the well.

The transmission amplitude $F(E)$ is a function of the energy E.

It was derived for $E > 0$.

If we assume that it is an analytic function of E, then it has some interesting properties.

In particular, in the regions of $E < 0$ and real, $F(E)$ has poles (goes to infinity) at E values that correspond to the bound state energies of the well.

$F(E)$ being infinite corresponds to having a transmitted wave without having an incident wave.

This is exactly the condition to have a bound state, where the transmitted(and reflected) wave does not propagate, but instead fall off exponentially.

The poles of $F(E)$ occur when

$$\cos k_1 a = \frac{i}{2}\left(\frac{k}{k_1} + \frac{k_1}{k}\right)\sin k_1 a$$
$$k_1 \cot \frac{k_1 a}{2} = ik \text{ and } k_1 \tan \frac{k_1 a}{2} = -ik$$

which are the same transcendental equations we had earlier for the bound states of the corresponding finite square well.

This property can be used to find bound state energies after solving the trans- mission problem for a given potential.

The book discusses wave packets and time delays.

Delta-Function Potentials

We now consider the potential energy function

where

$$V(x) = A\delta(x-a)$$
$$\delta(x-a) = 0 \quad x \neq a$$
$$\int_{-\infty}^{\infty} f(x)\delta(x-a)\,dx = f(a)$$

and solve the corresponding Schrodinger equation

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x)$$

As we discussed earlier the wave function $\psi(x)$ is assumed to be continuous for physical reasons relating to the probability interpretation.

The derivative of the wave function, however, is not continuous at $x = a$ for this potential.

We can see this as follows.

We have

$$-\frac{\hbar^2}{2m}\int_{a-\varepsilon}^{a+\varepsilon}\frac{d^2\psi(x)}{dx^2}dx + A\int_{a-\varepsilon}^{a+\varepsilon}\delta(x-a)V(x)\psi(x)dx = E\int_{a-\varepsilon}^{a+\varepsilon}\psi(x)dx$$

In the limit $\varepsilon \to 0$, using the continuity of $\psi(x)$, we get

$$-\frac{\hbar^2}{2m}\left[\left.\frac{d\psi}{dx}\right|_{a+\varepsilon} - \left.\frac{d\psi}{dx}\right|_{a-\varepsilon}\right] = E\psi(a)\int_{a-\varepsilon}^{a+\varepsilon}dx - A\psi(a)$$

so that

$$\text{discontinuity}\left(\frac{d\psi}{dx}\right)_{x=a} = \Delta\left(\frac{d\psi}{dx}\right) = \frac{2mA}{\hbar^2}\psi(a)$$

For simplicity we choose $a = 0$.

We then have two regions to consider

region I $\quad x < 0$

region II $\quad x > 0$

and the derivative is discontinuous at $x = 0$.

Transmission Problem

We first carry out the calculation of the transmission and reflection probabilities.

We assume that $A > 0$ (we have a delta function barrier), $E > 0$ and an incident wave of unit intensity coming in from the left.

In region I we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_I}{dx^2} = E\psi_I \rightarrow \psi_I(x) = e^{ikx} + Be^{-ikx} \text{ with } E = \frac{\hbar^2 k^2}{2m} > 0$$

We have both an incident and a reflected wave.

In region II we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{II}}{dx^2} = E\psi_{II} \rightarrow \psi_{II}(x) = Ce^{ikx} \text{ with } E = \frac{\hbar^2 k^2}{2m} > 0$$

There is only a transmitted wave.

The boundary conditions (at $x = 0$) give

$$\psi_I(0) = \psi_{II}(0) \rightarrow 1 + B = C$$

$$\frac{d\psi_{II}(0)}{dx} - \frac{d\psi_I(0)}{dx} = \frac{2m}{\hbar^2}A\psi_{II}(0) \rightarrow ikC - ik(1 - B) = \frac{2m}{\hbar^2}AC$$

The solutions are

$$C = \frac{ik}{ik - \frac{mA}{\hbar^2}} \text{ and } B = \frac{\frac{mA}{\hbar^2}}{ik - \frac{mA}{\hbar^2}}$$

We then have

$$T = \text{ transmission probability } = |C|^2 = \frac{1}{1 + \frac{mA^2}{2\hbar^2 E}}$$

$$R = \text{ reflection probability } = |B|^2 = \frac{1}{1 + \frac{2\hbar^2 E}{mA^2}}$$

We note that $T + R = 1$ as it must for the probability interpretation to make sense.

From our previous discussion, we suspect that the energy values of the poles of the transmission probability correspond to the bound state energies for the delta function well problem ($A < 0$).

For the single delta function potential, T has a single pole at

$$E = -\frac{mA^2}{2\hbar^2}$$

Bound-State Problem

We let $A \rightarrow -A,\ A > 0$.

In region I we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_I}{dx^2} = -|E|\,\psi_I \rightarrow \psi_I(x) = Be^{\alpha x}$$

with

$$E = -|E| = -\frac{\hbar^2\alpha^2}{2m} < 0$$

We have excluded the positive exponential term since it would diverge in region II as $x \rightarrow +\infty$.

The boundary conditions give

$$\psi_I(0) = \psi_{II}(0) \rightarrow B = C$$

$$\frac{d\psi_{II}(0)}{dx} - \frac{d\psi_I(0)}{dx} = -\frac{2m}{\hbar^2}A\psi_I(0) \rightarrow -\alpha C - \alpha B = -\frac{2m}{\hbar^2}AB$$

The resulting equation for α gives the allowed the bound state energies.

We have

$$\alpha = \frac{mA}{\hbar^2} \rightarrow \text{ only 1 solution only } \rightarrow \text{ 1 bound state}$$

$$E = -|E| = -\frac{\hbar^2\alpha^2}{2m} = -\frac{mA^2}{2\hbar^2}$$

which is the same value as we obtained from the pole of the transmission probability.

We also note that the solution has definite parity (even) since $\psi(x) = \psi(-x)$.

This must occur since $V(x) = V(-x)$ and hence parity commutes with the Hamiltonian.

As we also saw in the square well case, if only one solution exists then it is always an even parity solution.

Double Delta Function Potential

We now tackle a more realistic and hence more complex system, namely the double delta function potential energy function given by

$$V(x) = u\left[\delta\left(x + \frac{\ell}{2}\right) + \delta\left(x - \frac{\ell}{2}\right)\right]$$

We again start by considering the barrier transmission problem for $u > 0$.

Transmission problem

There are three regions to consider

region I	$x < \frac{\ell}{2}$
region II	$-\frac{\ell}{2} < x < \frac{\ell}{2}$
region III	$x > \frac{\ell}{2}$

The solutions in the three regions are

$$\begin{array}{lll} \text{region I} & x < \frac{\ell}{2} & \psi_I(x) = Ae^{ikx} + Be^{-ikx} \\ \text{region II} & -\frac{\ell}{2} < x < \frac{\ell}{2} & \psi_{II}(x) = Ce^{ikx} + De^{-ikx} \\ \text{region III} & x > \frac{\ell}{2} & \psi_{III}(x) = Fe^{ikx} \end{array} \qquad E = \frac{\hbar^2 k^2}{2m} > 0$$

The boundary conditions (the derivatives are discontinuous at $x = \pm \ell/2$)give the equations

$$x = -\frac{\ell}{2} \qquad Ae^{-ik\frac{\ell}{2}} + Be^{ik\frac{\ell}{2}} = Ce^{-ik\frac{\ell}{2}} + De^{ik\frac{\ell}{2}}$$

$$\text{ik}\left(Ce^{-ik\frac{\ell}{2}} - De^{ik\frac{\ell}{2}}\right) - ik\left(Ae^{-ik\frac{\ell}{2}} - Be^{ik\frac{\ell}{2}}\right)$$

$$= \frac{2mu}{\hbar^2}\left(Ce^{-ik\frac{\ell}{2}} + De^{ik\frac{\ell}{2}}\right)$$

$$x = +\frac{\ell}{2} \qquad Ce^{ik\frac{\ell}{2}} + De^{-ik\frac{\ell}{2}} = Fe^{ik\frac{\ell}{2}}$$

$$\text{ik}\left(Fe^{ik\frac{\ell}{2}}\right) - ik\left(Ce^{ik\frac{\ell}{2}} - De^{-ik\frac{\ell}{2}}\right) = \frac{2mu}{\hbar^2}\left(Fe^{ik\frac{\ell}{2}}\right)$$

We are interested in calculating the transmission probability

$$T(E) = \frac{|F|^2}{|A|^2}$$

Much algebra gives these results

$$C = (1 - \frac{mu}{ik\hbar^2})F \text{ and } D = \frac{mu}{ik\hbar^2}e^{ik\ell}F$$

$$\frac{F}{A} = \frac{1}{\left[1 - \frac{mu^2}{2\hbar^2 E}(1 - \cos 2k\ell)\right]^2 + i\left[\frac{ku}{E} + \frac{mu^2}{2\hbar^2 E}\sin 2k\ell\right]}$$

$$= \frac{1}{\left[1 - \frac{mu}{ik\hbar^2}\right]^2 - \left[\frac{mu}{ik\hbar^2}\right]^2 e^{2ik\ell}}$$

The poles of the transmission probability occur for the zeroes of the denominator.

If we let $u = -|u|$ and $E = -|E| < 0$ and $k = i\gamma$, the poles are given by

$$\left[1 - \frac{m|u|}{\gamma\hbar^2}\right]^2 - \left[\frac{m|u|}{\gamma\hbar^2}\right]^2 e^{-2\gamma\ell} = 0$$

Some rearranging and algebra gives

$$e^{\gamma\frac{\ell}{2}} = \frac{m|u|}{\gamma\hbar^2}(e^{\gamma\frac{\ell}{2}} \pm e^{-\gamma\frac{\ell}{2}})$$

as the transcendental equation for the bound state energies of the double delta function well.

We will solve this transcendental equation graphically.

Again, it is always best to first clean up the equation as much as possible by changing variables.

There are no rules for this part.

We let $$\beta = \gamma\frac{\ell}{2} \text{ and } \varepsilon = \frac{m|u|\ell}{2\hbar^2}$$

The transcendental equation then becomes

$$e^{-2\beta} = \pm\left(\frac{\beta}{\varepsilon} - 1\right) \qquad \pm \begin{matrix} \text{+1 parity} \\ \text{-1 parity} \end{matrix}$$

The book shows how to get the same result by actually solving the bound-state problem directly.

I will skip that.

Graphical Solution

Since

$$\varepsilon = \frac{m\,|u|\,\ell}{2\hbar^2}$$

is fixed by the choice of the parameters of the potential, we have transcendental equations for β.

The solutions give us the energy eigenvalues.

We can simplify the graphical solution with one more change of variables.

We let

$$\frac{\beta}{\varepsilon} = \eta$$

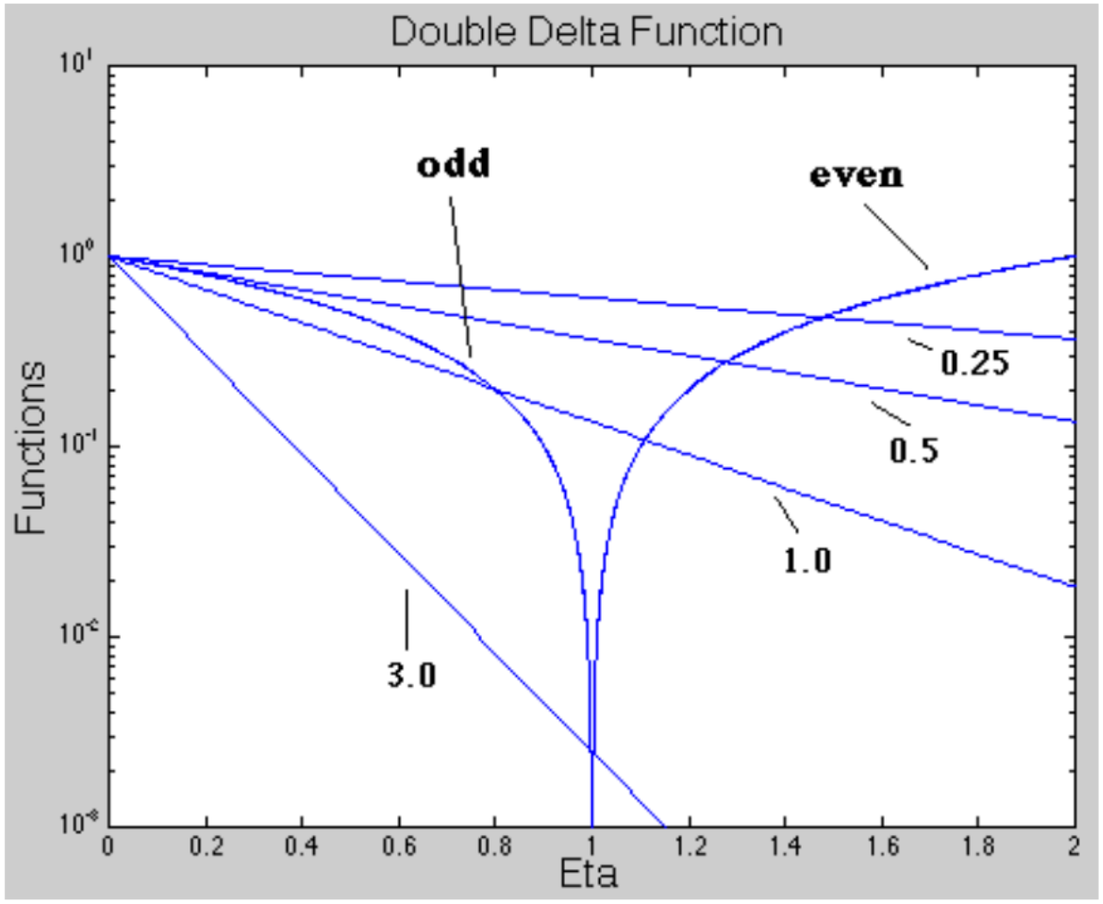
and get

$$\begin{array}{ll} \text{even parity} & e^{-2\epsilon\eta} = \eta - 1 \\ \text{odd parity} & e^{-2\epsilon\eta} = 1 - \eta \end{array}$$

Procedure:

1. plot $1-\eta$ versus η and $\eta - 1$ versus η
2. plot $e^{-2\epsilon\eta}$ versus η for several values of ϵ
3. the intersections are the even and odd solutions

In the Figure on the next slide, we plot the exponential function for several values of ε.



Double delta Function Bound States

As can be seen, the properties of the solutions are:

1. Even solutions always exist
2. Odd solutions do not always exists (see the $\epsilon = 0.25$ line, which only intersects the even curve)
3. As $\epsilon \to 0$ which corresponds to $\ell \to \infty$ we get $E_{odd} = E_{even}$. Physically, the wells are separating so that they no longer influence each other and the system behaves like two separate wells.

We will look at this property again later when we consider a translationally invariant infinite line of delta functions.

The interaction between the wells coupled to the translations invariance will lead to energy bands as in real solids.

Harmonic Oscillators

We now turn our attention to a physical system with a harmonic oscillator potential energy function

$$V(x) = \frac{1}{2}kx^2 = \frac{1}{2}m\omega_0^2 x^2$$

We will first solve this system by going to the position representation and solving the Schrodinger equation using differential equation techniques.

We will review all of our methods, introduce Hermite polynomials and define generating functions along the way.

After that we will introduce an algebraic method for solving this system that does not make any reference to the position representation.

Later, we will generalize this algebraic method for use in other types of systems.

Differential Equation Method

The Hamiltonian for the system is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega_0^2\hat{x}^2$$

The energy eigenvalue equation is

$$\hat{H}\left|\psi\right\rangle = E\left|\psi\right\rangle$$

We put this into the position representation by these steps:

$$\langle x|\,\hat{H}\,|\psi\rangle = \langle x|\,E\,|\psi\rangle = E\,\langle x\mid\psi\rangle = E\psi(x)$$

$$\langle x|\left(\frac{\hat{p}^2}{2m}+\frac{1}{2}m\omega_0^2\hat{x}^2\right)|\psi\rangle = E\psi(x)$$

$$\langle x|\,\frac{\hat{p}^2}{2m}\,|\psi\rangle + \frac{1}{2}m\omega_0^2\,\langle x|\,\hat{x}^2\,|\psi\rangle = E\psi(x)$$

$$\frac{1}{2m}(-i\hbar\frac{d}{dx})^2\,\langle x\mid\psi\rangle + \frac{1}{2}m\omega_0^2x^2\,\langle x\mid\psi\rangle = E\psi(x)$$

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2}+\frac{1}{2}m\omega_0^2x^2\psi(x) = E\psi(x)$$

The last line is the standard 1-dimensional time-independent Schrodinger equation for the harmonic oscillator potential.

When solving such equations it is useful to rewrite it in dimensionless form.

We do this by making these substitutions (the choice is an art, not a science)

$$\alpha^4 = \frac{m^2\omega_0^2}{\hbar^2}\ ,\ q{=}\alpha x \text{ and } \lambda = \frac{2E}{\hbar\omega_0}$$

The harmonic oscillator Schrodinger equation becomes

$$\frac{d^2\psi}{dq^2}+(\lambda-q^2)\psi = 0$$

The solution of such equations is greatly aided by examining and extracting the behavior of ψ in the asymptotic regions $q \to \pm\infty$.

For sufficiently large q the equation becomes

$$\frac{d^2\psi}{dq^2} - q^2\psi = 0$$

which is satisfied by functions of the form

$$q^n e^{\pm\frac{1}{2}q^2} \quad \text{for any finite value of } n$$

Since the probability interpretation of the function requires us to be able to normalize it, i.e., we must have

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

This rules out the positive exponential.

This result suggests that we try a solution of the form

$$\psi(q) = H(q)e^{-\frac{1}{2}q^2}$$

where $H(q)$ is a polynomial of finite order in q and we have explicitly extracted the asymptotic behavior.

Substituting into the original differential equation gives a new differential equation for $H(q)$

$$\frac{d^2H}{dq^2} - 2q\frac{dH}{dq} + (\lambda - 1)H = 0$$

We now use the standard Frobenius series substitution method to solve this equation.

We assume a solution of the form

$$H(q) = q^s \sum_{j=0}^{\infty} a_j q^j \text{ where } a_0 \neq 0 \text{ and } s \geq 0$$

The nonnegative restriction on s is required in order to guarantee that the solution is well-behaved at $q = 0$ (we need to be able to normalize the wave function).

If we substitute this guess into the differential equation for H we get

$$q^s \sum_{j=0}^{\infty} a_j (j+s)(j+s-1) q^{j-2} - 2q^s \sum_{j=0}^{\infty} a_j (j+s) q^j + (\lambda - 1) q^s \sum_{j=0}^{\infty} a_j q^j = 0$$

$$\begin{aligned} q^s \left[a_0 s(s-1) q^{-2} + a_1 s(s+1) q^{-1} + \left(a_2 (s+2)(s+1) - a_0 (2s+1-\lambda) \right) q^0 \right] \\ + \left(a_3 (s+3)(s+2) - a_1 (2s+3-\lambda) \right) q^1 + \ldots\ldots + \\ \left(a_{j+2} (s+j+2)(s+j+1) - (2s+2j+1-\lambda) a_j \right) q^j + \ldots\ldots = 0 \end{aligned}$$

Now for a power series to be identically zero, the coefficient of each power must be zero separately.

We thus have

$$\begin{aligned} s(s-1)a_0 &= 0 \\ s(s+1)a_1 &= 0 \\ (s+2)(s+1)a_2 - (2s+1-\lambda)a_0 &= 0 \\ (s+3)(s+2)a_3 - (2s+3-\lambda)a_1 &= 0 \\ &\ldots\ldots \\ (s+j+2)(s+j+1)a_{j+2} - (2s+2j+1-\lambda)a_j &= 0 \end{aligned}$$

Now we assumed that $a_0 \neq 0$, $s \geq 0$.

Therefore the equations says that we must have $s = 0$ or $s = 1$ and that $s = 0$ or $a_1 = 0$, or both.

The remaining equations give a_2 in terms of a_0, a_3 in terms of a_1, and, in general, a_{j+2} in terms of a_j.

Whether the series will have a finite or an infinite number of terms depends on the choice of s, a_1, and the eigenvalue λ.

If the series does not terminate, then its asymptotic behavior can be inferred from the coefficients of the high power terms.

From the last recursion formula we have

$$\frac{a_{j+2}}{a_j} \underset{j\to\infty}{\to} \frac{2}{j}$$

This ratio is the same as that of the series for $q^n e^{q^2}$ with any finite value of n.

This implies that the solution we have found for the wave function

$$\psi(q) = H(q)e^{-\frac{1}{2}q^2}$$

will diverge at large q.

Since we cannot have that happen and still retain the probability interpretation, the series must terminate ($H(q)$ must be a polynomial as I mentioned earlier).

From the last recursion formula we can see that if we choose

$$\lambda = 2s + 2j + 1$$

for some j, then the series will terminate ($a_{j+2} = 0$).

Since $a_0 \neq 0$ this cutoff procedure requires that j equals an even integer.

This means that the odd series will not terminate unless we choose $a_1 = 0$.

The index s can still be either 0 or 1 .

Corresponding to these two values, λ is equal to $2j+1$ or $2j+3$, where j is an even integer.

We can express both cases in terms of an integer n as

$$\lambda_n = 2n + 1$$

This gives us the energy eigenvalues

$$E_n = \frac{\lambda_n}{2}\hbar\omega_0 = \left(n + \frac{1}{2}\right)\hbar\omega_0$$

As in the square well examples, we have a *zero point energy*.

It corresponds to the lowest or ground state energy value

$$E_0 = \frac{1}{2}\hbar\omega_0$$

The eigenfunctions have a definite parity, as they should since $V(x) = V(-x)$.

From the definition of n, we can see that for a given solution

$$n = \text{ highest value of } s + j$$

If we denote the corresponding polynomials by $H_n(q)$ we see that $H_n(q)$ is of degree n in q.

Therefore, since the exponential is an even function, we have

n = even → even parity solution

n = odd → odd parity solution

The polynomial of order n, $H_n(q)$, is a solution of the original equation with $\lambda = 2n + 1$

$$\frac{d^2H_n}{dq^2} - 2q\frac{dH_n}{dq} + 2nH_n = 0$$

It is called the Hermite polynomial.

We could find the exact forms of the polynomials from the recursion relations.

However, it more instructive to introduce the idea of a generating function.

We define

$$S(q,s) = e^{q^2-(s-q)^2} = e^{-s^2+2sq} = \sum_{n=0}^{\infty} G_n(q)\frac{s^n}{n!}$$

The functions $G_n(q)$ defined in this way are polynomials in q.

Now we have

$$\frac{\partial S}{\partial q} = 2se^{-s^2+2sq} = \sum_{n=0}^{\infty} G_n(q)\frac{2s^{n+1}}{n!} = \sum_{n=0}^{\infty} \frac{dG_n(q)}{dq}\frac{s^n}{n!}$$

$$\frac{\partial S}{\partial s} = (-2s+2q)e^{-s^2+2sq} = \sum_{n=0}^{\infty} G_n(q)\frac{(-2s+2q)s^n}{n!} = \sum_{n=0}^{\infty} G_n(q)\frac{s^{n-1}}{(n-1)!}$$

Equating equal powers of q in each of these equation we get

$$\frac{dG_n}{dq} = 2nG_{n-1} \text{ and } G_{n+1} = 2qG_n - 2nG_{n-1}$$

The lowest order differential equation, which involves only $G_n(q)$, that we can construct out of this pair of equation is

$$\frac{d^2G_n}{dq^2} = 2n\frac{dG_{n-1}}{dq}, \frac{dG_{n+1}}{dq} = 2(n+1)G_n$$

$$\frac{dG_{n+1}}{dq} = 2\mathrm{q}\frac{dG_n}{dq} + 2G_n - 2\mathrm{n}\frac{dG_{n-1}}{dq}$$

$$\frac{d^2G_n}{dq^2} = 2\mathrm{q}\frac{dG_n}{dq} + 2G_n - \frac{dG_{n+1}}{dq} = 2\mathrm{q}\frac{dG_n}{dq} + 2G_n - 2(n+1)G_n$$

$$\frac{d^2G_n}{dq^2} - 2\mathrm{q}\frac{dG_n}{dq} - 2nG_n = 0$$

which is the Hermite polynomial equation.

So $G_n(q) = H_n(q)$ and we have

$$S(q,s) = e^{q^2-(s-q)^2} = e^{-s^2+2sq} = \sum_{n=0}^{\infty} H_n(q)\frac{s^n}{n!}$$

We then have

$$\frac{\partial^n S(q,s)}{\partial s^n} = e^{q^2}\frac{\partial^n e^{-(s-q)^2}}{\partial s^n} = (-1)^n e^{q^2}\frac{\partial^n e^{-(s-q)^2}}{\partial q^n}$$

$$= \frac{\partial^n \sum_{m=0}^{\infty} H_m(q)\frac{s^m}{m!}}{\partial s^n} = \sum_{m=n}^{\infty} H_m(q)\frac{s^{m-n}}{(m-n)!}$$

where we have used the property

$$\frac{\partial f(s-q)}{\partial s} = -\frac{\partial f(s-q)}{\partial q}$$

If we set $s = 0$ in the equation above we get

$$(-1)^n e^{q^2} \frac{\partial^n e^{-q^2}}{\partial q^n} = H_n(q)$$

We can then easily generate the polynomials

$$H_0 = 1\,,\; H_1 = 2q\,,\; H_2 = 4q^2 - 2\,,\; \ldots$$

The Schrodinger wave functions are then given by

$$\psi_n(x) = A_n H_n(\alpha x) e^{-\frac{1}{2}\alpha^2 x^2}$$

We find the normalization constant A_n by

$$\int_{-\infty}^{\infty} |\psi_n(x)|^2\, dx = 1 = \frac{|A_n|^2}{\alpha} \int_{-\infty}^{\infty} H_n^2(q) e^{-q^2}\, dq$$

Now we can determine all of the integrals of this type from the generating function as follows.

We have

$$\int_{-\infty}^{\infty} e^{-s^2+2sq} e^{-t^2+2tq} e^{-q^2}\, dq = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{s^n t^m}{n!m!} \int_{-\infty}^{\infty} H_n(q) H_m(q) e^{-q^2}\, dq$$

$$= \sqrt{\pi} e^{2st} = \sqrt{\pi} \sum_{n=0}^{\infty} \frac{(2st)^n}{n!}$$

which says that

$$\int_{-\infty}^{\infty} H_n(q) H_m(q) e^{-q^2}\, dq = \sqrt{\pi} 2^n n! \delta_{nm}$$

The case $n = m$, gives

$$A_n = \left(\frac{\alpha}{\sqrt{\pi} 2^n n!} \right)^{1/2}$$

The case $n \neq m$ just corresponds to the orthogonality property of eigenfunctions of different eigenvalues.

Algebraic Method

The Hamiltonian for this system is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2} m \omega_0^2 \hat{x}^2$$

The energy eigenvalue equation is

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

The commutator between the position and momentum operators is

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

This information defines a complete solution to the problem when using algebraic methods.

We now define the two new operators (like a change of basis)

$$\hat{a} = \sqrt{\frac{m\omega_0}{2\hbar}} \hat{x} + \frac{i\hat{p}}{\sqrt{2m\hbar\omega_0}} \text{ and } \hat{a}^+ = \sqrt{\frac{m\omega_0}{2\hbar}} \hat{x} - \frac{i\hat{p}}{\sqrt{2m\hbar\omega_0}}$$

$$\hat{x} = \sqrt{\frac{\hbar}{2m\omega_0}} (\hat{a} + \hat{a}^+) \text{ and } \hat{p} = -i\sqrt{\frac{m\hbar\omega_0}{2}} (\hat{a} - \hat{a}^+)$$

These new operators are not Hermitian.

They are Hermitian conjugates, however.

We then have

$$[\hat{a}, \hat{a}^{+}] = 1 \text{ and } \hat{H} = \hbar\omega_0(\hat{N}_{op} + \frac{1}{2}), \hat{N}_{op} = \hat{a}^{+}\hat{a}$$

Now suppose that the vector $|n\rangle$ is an eigenvector of $\hat{N}_{op} = \hat{a}^{\dagger}\hat{a}$ with eigenvalue n

$$\hat{N}_{op}|n\rangle = n|n\rangle$$

Since $\hat{N}_{op}$ and $\hat{H}$ differ only by a constant, this implies that $|n\rangle$ is also an eigenvector of $\hat{H}$, i.e.,

$$\hat{H}|n\rangle = \hbar\omega_0(\hat{N}_{op} + \frac{1}{2})|n\rangle = \hbar\omega_0(n + \frac{1}{2})|n\rangle$$

Therefore, $|E_n\rangle = |n\rangle$ and the energy eigenvalues are

$$E_n = \hbar\omega_0(n + \frac{1}{2})$$

Our task now is to find the allowed values of n.

Now $\hat{N}_{op}^{\dagger} = (\hat{a}^{\dagger}\hat{a}) = \hat{a}^{\dagger}\hat{a} = \hat{N}_{op}$, so $\hat{N}_{op}$ is a Hermitian operator.

This means that the eigenvalues of $\hat{N}_{op}$ (and $\hat{H}$) are real numbers.

Now we also have the relation

$$n = \langle n|\hat{N}_{op}|n\rangle = \langle n|\hat{a}^{+}\hat{a}|n\rangle$$

If we let $|\phi\rangle = \hat{a}|n\rangle$, then the last equation can be written

$$n = \langle \phi \mid \phi \rangle \geq 0$$

which says that the eigenvalues of $\hat{N}_{op}$ (the values of n) are *non-negative*, *real* numbers.

Finally, using $[\hat{a}, \hat{a}^{\dagger}] = 1$, we have

$$\hat{N}_{op}\hat{a} = \hat{a}^{+}\hat{a}\hat{a} = (\hat{a}\hat{a}^{+} - 1)\hat{a} = \hat{a}(\hat{a}^{+}\hat{a} - 1) = \hat{a}\left(\hat{N}_{op} - 1\right)$$

and similarly,

$$\hat{N}_{op}\hat{a}^{+} = \hat{a}^{+}\left(\hat{N}_{op} + 1\right)$$

These operator relations imply that

$$\hat{N}_{op}\hat{a}\,|n\rangle = \hat{a}(\hat{N}_{op} - 1)\,|n\rangle = (n-1)\hat{a}\,|n\rangle$$

which says that $\hat{a}\,|n\rangle$ is an eigenstate of $\hat{N}_{op}$ with eigenvalue $n-1$ or that we can write

$$\hat{a}\,|n\rangle = \alpha\,|n-1\rangle \rightarrow \langle n|\,\hat{a}^{+} = \alpha^{*}\,\langle n-1|$$

We will assume that the eigenstates are normalized, i.e., $\langle n \mid n\rangle = 1$.

We then have

$$(\langle n|\,\hat{a}^{+})\,(\hat{a}\,|n\rangle) = |\alpha|^{2}\,\langle n-1 \mid n-1\rangle = |\alpha|^{2}$$

$$= \langle n|\,\hat{a}^{+}\hat{a}\,|n\rangle = \langle n|\,\hat{N}_{op}\,|n\rangle = n$$

which says that $\alpha = \sqrt{n}$ and we have the relation

$$\hat{a}\,|n\rangle = \sqrt{n}\,|n-1\rangle$$

Repeating this procedure we have

$$\hat{a}^{2}\,|n\rangle = \hat{a}\sqrt{n}\,|n-1\rangle = \sqrt{n}\hat{a}\,|n-1\rangle = \sqrt{n(n-1)}\,|n-2\rangle$$

which implies the following:

if the eigenvalue n exists, then
so do the eigenvalues $n-1$, $n-2$, ...

This sequence cannot continue forever since we have already shown that the eigenvalues of $\hat{N}_{op}$ are non-negative.

The only way out of this dilemma is that *n* is an integer.

In this case, we would eventually arrive at some eigenstate and then get

$$\hat{a}\,|1\rangle = |0\rangle \text{ and } \hat{a}\,|0\rangle = 0$$

which says the sequence ends.

Thus, the eigenvalues of $\hat{N}_{op}$ are the *non-negative integers*.

We can also go the other way since similar algebra shows that

$$\hat{N}_{op}\hat{a}^{+}\,|n\rangle = \hat{a}^{+}(\hat{N}_{op}+1)\,|n\rangle = (n+1)\hat{a}^{+}\,|n\rangle$$

This implies that, in the same way as before,

$$\hat{a}^{+}\,|n\rangle = \sqrt{n+1}\,|n+1\rangle$$

Thus, the eigenvalues of N_{op} are all the non-negative integers

$$n = 0, 1, 2, 3, 4, \ldots\ldots$$

The operators $\hat{a}$ and $\hat{a}^{\dagger}$ are called lowering/raising operators from their property of lowering and raising the eigenvalues by unity.

They are also sometimes called ladder operators because they generate a ladder of equally-spaced eigenvalues.

Finally, assuming that $\langle 0 \mid 0 \rangle = 1$ we have

$$|1\rangle = \frac{\hat{a}^+}{\sqrt{1}}|0\rangle = \frac{\hat{a}^+}{\sqrt{1!}}|0\rangle$$

$$|2\rangle = \frac{\hat{a}^+}{\sqrt{2}}|1\rangle = \frac{\hat{a}^+}{\sqrt{2}}\frac{\hat{a}^+}{\sqrt{1}}|0\rangle = \frac{(\hat{a}^+)^2}{\sqrt{2!}}|0\rangle$$

$$|3\rangle = \frac{\hat{a}^+}{\sqrt{3}}|2\rangle = \frac{\hat{a}^+}{\sqrt{3}}\frac{(\hat{a}^+)^2}{\sqrt{2!}}|0\rangle = \frac{(\hat{a}^+)^3}{\sqrt{3!}}|0\rangle$$

.....

$$|n\rangle = \frac{(\hat{a}^+)^n}{\sqrt{n!}}|0\rangle$$

and

$$E_n = \hbar\omega_0(n + \frac{1}{2})$$

Notice that we were able to derive all of these results without introducing the wave function (the position representation).

If you are desperate to see the position representation wave functions they can easily be derived from this formalism.

We have

$$0 = \langle x|\,\hat{a}\,|0\rangle = \langle x|\sqrt{\frac{m\omega_0}{2\hbar}}\hat{x} + \frac{i\hat{p}}{\sqrt{2m\hbar\omega_0}}|0\rangle$$

$$= \sqrt{\frac{m\omega_0}{2\hbar}}x\,\langle x\mid 0\rangle + \frac{i(-i\hbar\frac{d}{dx})}{\sqrt{2m\hbar\omega_0}}\langle x\mid 0\rangle$$

This is a differential equation for $\langle x \mid 0\rangle = \psi_0(x)$ which is the ground state wave function.

We have

$$\frac{d\psi_o(x)}{dx} + \frac{m\omega_0}{\hbar} x\psi_o(x) = 0$$

which has the solution (when normalized)

$$\langle x \mid 0\rangle = \psi_o(x) = \left(\frac{m\omega_0}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega_0 x^2}{2\hbar}}$$

which agrees with our earlier work.

The higher energy wave functions are obtained using

$$\begin{aligned}\langle x \mid 1\rangle = \psi_1(x) &= \langle x|\, \hat{a}^+ \,|0\rangle \\ &= \langle x| \sqrt{\frac{m\omega_0}{2\hbar}}\hat{x} - \frac{i\hat{p}}{\sqrt{2m\hbar\omega_0}} \,|0\rangle \\ &= \sqrt{\frac{2m\omega_0}{\hbar}} x \,\langle x \mid 0\rangle\end{aligned}$$

and so on.

The raising and lowering operator formalism is very powerful and will be very useful beyond non-relativistic quantum mechanics where they will allow the theory to represent the creation and annihilation of particles.

Coherent States

Let us now use this formalism to do an interesting exercise and show the power of the methods.

We consider the following question: what are the eigenstates of the lowering operator $\hat{a}$?

That is, we write

$$\hat{a}\,|\alpha\rangle = \alpha\,|\alpha\rangle \;\text{ with }\; \alpha = |\alpha|\, e^{i\phi}$$

where $|\alpha\rangle$ is the eigenvector of $\hat{a}$ and α is the eigenvalue, which is not necessarily real since $\hat{a}$ is not Hermitian.

Since the vectors $|n\rangle$ are eigenvectors of a Hermitian operator, they form a orthonormal complete set and can be used as an orthonormal basis vector space.

We can then write

$$|\alpha\rangle = \sum_{m=0}^{\infty} b_m\,|m\rangle$$

where

$$\langle k \mid \alpha\rangle = \sum_{m=0}^{\infty} b_m \langle k \mid m\rangle = \sum_{m=0}^{\infty} b_m \delta_{km} = b_k$$

Now

$$\langle n-1|\,\hat{a}\,|\alpha\rangle = \alpha\,\langle n-1 \mid \alpha\rangle = \alpha b_{n-1}$$

and using

$$\hat{a}^{+}\,|n-1\rangle = \sqrt{n}\,|n\rangle \rightarrow \langle n-1|\,\hat{a} = \sqrt{n}\,\langle n|$$

we have

$$\langle n-1|\,\hat{a}\,|\alpha\rangle = \sqrt{n}\,\langle n \mid \alpha\rangle = \sqrt{n} b_n$$

or

$$b_n = \frac{\alpha}{\sqrt{n}} b_{n-1}$$

This says that

$$b_1 = \frac{\alpha}{\sqrt{1}} b_0 \;,\; b_2 = \frac{\alpha}{\sqrt{2}} b_1 = \frac{\alpha^2}{\sqrt{2!}} b_0$$

or

$$b_n = \frac{\alpha^n}{\sqrt{n!}} b_0$$

We thus get the final result

$$|\alpha\rangle = b_0 \sum_{m=0}^{\infty} \frac{\alpha^m}{\sqrt{m!}} |m\rangle$$

Let us now normalize this state (choose b_0). We have

$$\begin{aligned}
\langle \alpha \mid \alpha \rangle = 1 &= |b_0|^2 \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{\alpha^{*m} \alpha^k}{\sqrt{m!}\sqrt{k!}} \langle k \mid m \rangle \\
&= |b_0|^2 \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{\alpha^{*m} \alpha^k}{\sqrt{m!}\sqrt{k!}} \delta_{km} \\
&= |b_0|^2 \sum_{m=0}^{\infty} \frac{|\alpha|^2}{m!} = |b_0|^2 e^{|\alpha|^2}
\end{aligned}$$

which says that

$$b_0 = e^{-\frac{1}{2}|\alpha|^2}$$

and thus

$$|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{m=0}^{\infty} \frac{\alpha^m}{\sqrt{m!}} |m\rangle$$

Now

$$\begin{aligned}
\langle n \mid \alpha \rangle = &\text{ probability amplitude that the system in the state} \\
&|\alpha\rangle \text{ will be found in the state } |n\rangle
\end{aligned}$$

which then says that

$$P_n = |\langle n \mid \alpha \rangle|^2 = \frac{e^{-|\alpha|^2} |\alpha|^{2n}}{n!} = \frac{e^{-N} N^n}{n!}$$

is the probability that the system in the state $|\alpha\rangle$ will be found in the state $|n\rangle$, where we have defined $N = |\alpha|^2$.

We note that

$$\langle\alpha|\,\hat{a}^{+}\hat{a}\,|\alpha\rangle = \left|\alpha^{2}\right|\langle\alpha\mid\alpha\rangle = \left|\alpha^{2}\right| = N = \langle\alpha|\,\hat{N}_{op}\,|\alpha\rangle$$

or N = the average value or expectation value of the $\hat{N}_{op}$ operator in the state $|\alpha\rangle$.

This type of probability distribution is called a Poisson distribution, i.e., the state $|\alpha\rangle$ has the number states or energy eigenstates distributed in a Poisson manner.

Since the states $|n\rangle$ are energy eigenstates, we know their time dependence, i.e.,

$$|n,t\rangle = e^{-i\frac{E_n}{\hbar}t}\,|n\rangle$$

Therefore, we have for the time dependence of the state $|\alpha\rangle$

$$|\alpha,t\rangle = e^{-\frac{1}{2}|\alpha|^2}\sum_{m=0}^{\infty}\frac{\alpha^m}{\sqrt{m!}}\,|m,t\rangle = e^{-\frac{1}{2}|\alpha|^2}\sum_{m=0}^{\infty}\frac{\alpha^m}{\sqrt{m!}}e^{-i\frac{E_m}{\hbar}t}\,|m\rangle$$

This simple operation clearly indicates the fundamental importance of the energy eigenstates when used as a basis set.

If we are able to expand an arbitrary vector representing some physical system in the energy basis, then we immediately know the time dependence of that state vector and hence we know the time dependence of all the probabilities associated with the state vector and the system.

We will use this result to solve for the time dependence of real physical systems in later discussions.

Now let us try to understand the physics contained in the $|\alpha\rangle$ state vector.

In a given energy eigenstate the expectation value of the position operator is given by

$$\langle n,t|\,\hat{x}\,|n,t\rangle = \sqrt{\frac{\hbar}{2m\omega_0}}\,\langle n,t|\,(\hat{a}+\hat{a}^+)\,|n,t\rangle$$
$$= \sqrt{\frac{\hbar}{2m\omega_0}}\,\langle n|\,e^{i\frac{E_n}{\hbar}t}(\hat{a}+\hat{a}^+)e^{-i\frac{E_n}{\hbar}t}\,|n\rangle$$
$$= \sqrt{\frac{\hbar}{2m\omega_0}}\,\langle n|\,(\hat{a}+\hat{a}^+)\,|n\rangle$$
$$= \sqrt{\frac{\hbar}{2m\omega_0}}\,\langle n|\left(\sqrt{n}\,|n-1\rangle+\sqrt{n+1}\,|n+1\rangle\right)=0$$

i.e., it is equal to zero and is independent of time.

On the other hand, in the state $|\alpha\rangle$ we find

$$\langle \alpha,t|\,\hat{x}\,|\alpha,t\rangle = \sqrt{\frac{\hbar}{2m\omega_0}}\sum_m\sum_k b_m^* b_k e^{i\frac{(E_m-E_k)}{\hbar}t}\,\langle m|\,(\hat{a}+\hat{a}^+)\,|k\rangle$$

Now

$$\langle m|\,(\hat{a}+\hat{a}^+)\,|k\rangle = \langle m|\left(\sqrt{k}\,|k-1\rangle+\sqrt{k+1}\,|k+1\rangle\right)$$
$$= \sqrt{k}\delta_{m,k-1}+\sqrt{k+1}\delta_{m,k+1}$$

Using this result we have

$$
\begin{aligned}
&\langle \alpha, t| \hat{x} |\alpha, t\rangle \\
&\quad = \sqrt{\frac{\hbar}{2m\omega_0}} \left(\sum_{k=1}^{\infty} b_{k-1}^* b_k \sqrt{k} e^{i\frac{(E_{k-1}-E_k)}{\hbar}t} + \sum_{k=0}^{\infty} b_{k+1}^* b_k \sqrt{k+1} e^{i\frac{(E_{k+1}-E_k)}{\hbar}t} \right) \\
&\quad = \sqrt{\frac{\hbar}{2m\omega_0}} \left(\sum_{k=1}^{\infty} b_{k-1}^* b_k \sqrt{k} e^{-i\omega_0 t} + \sum_{k=0}^{\infty} b_{k+1}^* b_k \sqrt{k+1} e^{i\omega_0 t} \right) \\
&\quad = \sqrt{\frac{\hbar}{2m\omega_0}} \left(\sum_{k=0}^{\infty} b_k^* b_{k+1} \sqrt{k} e^{-i\omega_0 t} + \sum_{k=0}^{\infty} b_{k+1}^* b_k \sqrt{k+1} e^{i\omega_0 t} \right) \\
&\quad = \sqrt{\frac{\hbar}{2m\omega_0}} b_0^2 \left(\sum_{k=0}^{\infty} \frac{\alpha^{*k}\alpha^{k+1}}{\sqrt{(k+1)!k!}} \sqrt{k} e^{-i\omega_0 t} + \sum_{k=0}^{\infty} \frac{\alpha^{*k+1}\alpha^{k}}{\sqrt{(k+1)!k!}} \sqrt{k+1} e^{i\omega_0 t} \right) \\
&\quad = \sqrt{\frac{\hbar}{2m\omega_0}} b_0^2 \sum_k \frac{1}{k!} |\alpha|^{2k} \left(\alpha e^{-i\omega_0 t} + \alpha^* e^{i\omega_0 t} \right)
\end{aligned}
$$

Now using $\alpha = |\alpha| e^{i\phi}$ we get

$$
\begin{aligned}
\langle \alpha, t| \hat{x} |\alpha, t\rangle &= \sqrt{\frac{\hbar}{2m\omega_0}} b_0^2 2 |\alpha| \sum_k \frac{|\alpha|^{2k}}{k!} \mathrm{Real}(e^{i\phi} e^{-i\omega_0 t}) \\
&= 2x_0 |\alpha| \cos(\omega_0 t - \phi) (b_0^2 \sum_k \frac{|\alpha|^{2k}}{k!}) \quad , \quad x_0 = \sqrt{\frac{\hbar}{2m\omega_0}} \\
&= 2x_0 |\alpha| \cos(\omega_0 t - \phi)
\end{aligned}
$$

The expectation value of the position operator in the state $|\alpha\rangle$ is time-dependent and behaves like a classical oscillator.

Now let us look at these states in another way.

We consider adding a term to the Hamiltonian of the form

$$V(\hat{x}) = -F_0\hat{x}$$

so that

$$\hat{H} = \hbar\omega_0(\hat{a}^+\hat{a} + \frac{1}{2}) - F_0x_0(\hat{a} + \hat{a}^+)$$

This corresponds to putting a charged oscillator in an electric field and including the effective dipole interaction energy in the Hamiltonian.

In the corresponding classical mechanics problem, we have

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2 \rightarrow \text{ oscillator with equilibrium point at } x = 0$$

and if we add a linear term in x, we get

$$\begin{aligned} H &= \frac{p^2}{2m} + \frac{1}{2}kx^2 - F_0x \\ &= \frac{p^2}{2m} + \frac{1}{2}k(x - \frac{F_0}{k})^2 - \frac{F_0^2}{2k} \\ &\rightarrow \text{ oscillator with equilibrium point at } x = \frac{F_0}{k} \end{aligned}$$

This suggests that we look for an oscillator solution to the quantum mechanical problem.

In particular, we look for a new set of raising and lowering operators with the same commutator as the original set.

We can accomplish this by letting

$$\hat{A} = \hat{a} + \beta \text{ and } \hat{A}^{+} = \hat{a}^{+} + \beta \text{ where } \beta = \text{ a real number}$$

The commutator $[\hat{a}, \hat{a}^{\dagger}] = 1$ then implies that $[\hat{A}, \hat{A}^{\dagger}] = 1$.

The Hamiltonian becomes

$$\hat{H} = \hbar\omega_0((\hat{A}^{+} - \beta)(\hat{A} - \beta) + \frac{1}{2}) - F_0 x_0((\hat{A} - \beta) + (\hat{A}^{+} - \beta))$$

$$= \hbar\omega_0(\hat{A}^{+}\hat{A} + \frac{1}{2}) - \hbar\omega_0\beta(\hat{A} + \hat{A}^{+}) + \hbar\omega_0\beta^2 - F_0 x_0(\hat{A} + \hat{A}^{+}) + 2F_0 x_0\beta$$

We are still free to choose β.

We choose $\omega_0\beta = -F_0 x_0$, which gives

$$\hat{H} = \hbar\omega_0(\hat{A}^{+}\hat{A} + \frac{1}{2}) - \frac{F_0^2 x_0^2}{\hbar\omega_0}$$

which is just a harmonic oscillator with the energies shifted by a constant.

Now choose the new number states $|N\rangle$ where $\hat{A}^{\dagger}\hat{A}\,|N\rangle = N\,|N\rangle$ as before.

We then have

$$\hat{H}\,|N\rangle = (\hbar\omega_0(\hat{A}^{+}\hat{A} + \frac{1}{2}) - \frac{F_0^2 x_0^2}{\hbar\omega_0})\,|N\rangle$$

$$= (\hbar\omega_0(N + \frac{1}{2}) - \frac{F_0^2 x_0^2}{\hbar\omega_0})\,|N\rangle = E_N\,|N\rangle$$

or we get the same set of energy levels as before, except that they are all displaced by the same constant amount

$$-\frac{F_0^2 x_0^2}{\hbar\omega_0}$$

The new ground state is given by

$$\hat{A}\,|N=0\rangle = (\hat{a}+\beta)\,|N=0\rangle = 0$$

$$\hat{a}\,|N=0\rangle = -\beta\,|N=0\rangle = \frac{F_0 x_0}{\hbar\omega_0}\,|N=0\rangle$$

The new ground state is just an eigenstate of $\hat{a}$ with eigenvalue

$$\alpha = \frac{F_0 x_0}{\hbar\omega_0}$$

i.e., the new ground state is a *coherent state*.

These states are called *coherent* states because of their properties when they appear in the quantum mechanical theory of the laser.

We will discuss these states in many other contexts later.

Now let us get this state another way that illustrates some new powerful tools.

Using the Translation Operator

In general, a displaced state $|\lambda\rangle$ is given in terms of the displacement operator (in one dimension) by

$$|\lambda\rangle = e^{-\frac{i}{\hbar}\hat{p}\lambda}\,|0\rangle$$

For the harmonic oscillator system

$$\hat{p} = \frac{1}{i}\sqrt{\frac{m\hbar\omega}{2}}(\hat{a}-\hat{a}^{+})$$

If we choose $|0\rangle$ to be the ground state of the oscillator, then we have for the corresponding displaced ground-state

$$|\lambda\rangle = e^{\sqrt{\frac{m\omega}{2\hbar}}(\hat{a}^+ - \hat{a})\lambda}\,|0\rangle$$

By Glauber's theorem (see proof below)

$$e^{(\hat{A}+\hat{B})} = e^{\hat{A}} e^{\hat{B}} e^{-\frac{1}{2}[\hat{A},\hat{B}]}$$

we have

$$e^{\sqrt{\frac{m\omega}{2\hbar}}(\hat{a}^+ - \hat{a})\lambda} = e^{\sqrt{\frac{m\omega}{2\hbar}}\hat{a}^+\lambda} e^{-\sqrt{\frac{m\omega}{2\hbar}}\hat{a}\lambda} e^{\frac{1}{2}\frac{m\omega}{2\hbar}[\hat{a}^+,\hat{a}]\lambda^2}$$

$$= e^{\sqrt{\frac{m\omega}{2\hbar}}\hat{a}^+\lambda} e^{-\sqrt{\frac{m\omega}{2\hbar}}\hat{a}\lambda} e^{-\frac{1}{4}\frac{m\omega}{\hbar}\lambda^2}$$

and thus

$$|\lambda\rangle = e^{\sqrt{\frac{m\omega}{2\hbar}}\hat{a}^+\lambda} e^{-\sqrt{\frac{m\omega}{2\hbar}}\hat{a}\lambda} e^{-\frac{1}{4}\frac{m\omega}{\hbar}\lambda^2}\,|0\rangle$$

Now

$$e^{-\sqrt{\frac{m\omega}{2\hbar}}\hat{a}\lambda}\,|0\rangle = \left(\hat{I} + \left(-\sqrt{\frac{m\omega}{2\hbar}}\lambda\hat{a}\right) + \frac{1}{2}\left(-\sqrt{\frac{m\omega}{2\hbar}}\lambda\hat{a}\right)^2 + \ldots\right)|0\rangle$$

$$= |0\rangle$$

where we used $\hat{a}\,|0\rangle = 0$.

Similarly, using $(\hat{a}^\dagger)^n\,|0\rangle = \sqrt{n!}\,|n\rangle$ we have

$$e^{\sqrt{\frac{m\omega}{2\hbar}}\hat{a}^{+}\lambda}|0\rangle = \left(\hat{I} + \left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\hat{a}^{+}\right) + \frac{1}{2}\left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\hat{a}^{+}\right)^{2} + \ldots\right)|0\rangle$$

$$= |0\rangle + \sqrt{\frac{m\omega}{2\hbar}}\lambda|1\rangle + \frac{1}{2}\left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\right)^{2}|2\rangle + \ldots$$

$$= \sum_{n=0}^{\infty} \frac{\left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\right)^{n}}{\sqrt{n!}}|n\rangle$$

or

$$|\lambda\rangle = e^{-\frac{1}{4}\frac{m\omega}{\hbar}\lambda^{2}} \sum_{n=0}^{\infty} \frac{\left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\right)^{n}}{\sqrt{n!}}|n\rangle$$

Thus

$$|\lambda\rangle = \sum_{n=0}^{\infty} b_n |n\rangle$$

where

$$b_n = \frac{e^{-\frac{N}{2}} N^{\frac{n}{2}}}{\sqrt{n!}} \quad , \quad \frac{N}{2} = \frac{m\omega}{4\hbar}\lambda^{2}$$

or

$$P_n = \text{ probability of find the system in the state } |n\rangle$$

$$= |b_n|^{2} = \frac{e^{-N} N^{n}}{n!}$$

which is a Poisson distribution.

We, thus, obtain the coherent states once again.

Glauber's theorem proof:

For two operators $\hat{A}$ and $\hat{B}$ assume that

$$\left[(\hat{A}+\hat{B}),\left[\hat{A},\hat{B}\right]\right] = \left[\hat{A},\left[\hat{A},\hat{B}\right]\right] = \left[\hat{B},\left[\hat{A},\hat{B}\right]\right] = 0 \qquad \text{(proof in book)}$$

Now let x be a parameter and consider the operator

$$f(x) = e^{\hat{A}x}e^{\hat{B}x}$$

We then get

$$\frac{df}{dx} = \hat{A}e^{\hat{A}x}e^{\hat{B}x} + e^{\hat{A}x}\hat{B}e^{\hat{B}x} = \hat{A}e^{\hat{A}x}e^{\hat{B}x} + e^{\hat{A}x}\hat{B}e^{-\hat{A}x}e^{\hat{A}x}e^{\hat{B}x}$$

$$= \left(\hat{A} + e^{\hat{A}x}\hat{B}e^{-\hat{A}x}\right)e^{\hat{A}x}e^{\hat{B}x} = \left(\hat{A} + e^{\hat{A}x}\hat{B}e^{-\hat{A}x}\right)f(x)$$

Our assumption above implies that

$$\left[\hat{B},\hat{A}^n\right] = n\hat{A}^{n-1}\left[\hat{B},\hat{A}\right]$$

and

$$\left[\hat{B},e^{-\hat{A}x}\right] = \sum_n (-1)^n \frac{x^n}{n!}\left[\hat{B},\hat{A}^n\right]$$

$$= \sum_n (-1)^n \frac{x^n}{(n-1)!}\hat{A}^{n-1}\left[\hat{B},\hat{A}\right]$$

$$= -e^{-\hat{A}x}\left[\hat{B},\hat{A}\right]x$$

which gives

$$\left[\hat{B},e^{-\hat{A}x}\right] = \hat{B}e^{-\hat{A}x} - e^{-\hat{A}x}\hat{B} = -e^{-\hat{A}x}\left[\hat{B},\hat{A}\right]x$$

$$e^{\hat{A}x}\hat{B}e^{-\hat{A}x} = \hat{B} - \left[\hat{B},\hat{A}\right]x$$

Thus, we have

$$\frac{df}{dx} = \left(\hat{A} + e^{\hat{A}x}\hat{B}e^{-\hat{A}x}\right) f(x) = \left(\hat{A} + \hat{B} + \left[\hat{A}, \hat{B}\right] x\right) f(x)$$

Since $\left[(\hat{A} + \hat{B}), \left[\hat{A}, \hat{B}\right]\right] = 0$, we can treat them like numbers to solve the above differential equation, i.e., we do not have to be careful above operator order.

The solution is

$$f(x) = e^{\hat{A}x} e^{\hat{B}x} = e^{(\hat{A}+\hat{B})x} e^{\frac{1}{2}\left[\hat{A},\hat{B}\right]x^2}$$

Letting $x = 1$ we have the Baker-Hausdorf identity (Glauber's theorem)

$$e^{(\hat{A}+\hat{B})} = e^{\hat{A}} e^{\hat{B}} e^{-\frac{1}{2}\left[\hat{A},\hat{B}\right]}$$

Charged Particles in Electromagnetic Fields (Uses Classical Mechanics)

If a charged particle moves in an electromagnetic field, the Lorentz force

$$\vec{F} = q\left(\vec{E} + \frac{\vec{v}}{c} \times \vec{B}\right)$$

acts on the particle, where q = charge, $\vec{E}$ = electric field, $\vec{\mathrm{v}}$ = velocity and $\vec{B}$ = magnetic field.

The electric and magnetic fields can be expressed in terms of the electromagnetic vector and scalar potentials

$$\vec{E} = -\nabla\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t} \text{ and } \vec{B} = \nabla \times \vec{A}$$

In order to make the transition to a quantum mechanical discussion, we must determine the Lagrangian and Hamiltonian for the charged particle in an electromagnetic field.

The appearance of the velocity vector $\vec{v}$ in the Lorentz force law means that we will need to work with a velocity-dependent potential function.

If we insert the expressions for the electric and magnetic fields into the Lorentz force law we get

$$\vec{F} = q\left(-\nabla\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t} + \frac{1}{c}\vec{v} \times \left(\nabla \times \vec{A}\right)\right)$$

Using the vector identity

$$\vec{B} \times \left(\nabla \times \vec{C}\right) = \nabla\left(\vec{B} \cdot \vec{C}\right) - \left(\vec{B} \cdot \nabla\right)\vec{C} - \left(\vec{C} \cdot \nabla\right)\vec{B} - \vec{C} \times \left(\nabla \times \vec{B}\right)$$

we get

$$\vec{v} \times \left(\nabla \times \vec{A}\right) = \nabla\left(\vec{v} \cdot \vec{A}\right) - \left(\vec{v} \cdot \nabla\right)\vec{A}$$

where we have used the fact that $\vec{v}$ is not an explicit function of the position vector $\vec{r}$.

Now, the total derivative relationship is given by

$$\frac{d\vec{A}}{dt} = \frac{\partial \vec{A}}{\partial t} + \frac{\partial \vec{A}}{\partial x}\frac{dx}{dt} + \frac{\partial \vec{A}}{\partial y}\frac{dy}{dt} + \frac{\partial \vec{A}}{\partial z}\frac{dz}{dt}$$

$$= \frac{\partial \vec{A}}{\partial t} + (\vec{v} \cdot \nabla)\vec{A}$$

Putting all this together we get

$$\vec{F} = q\left(-\nabla\varphi + \frac{1}{c}\nabla(\vec{v} \cdot \vec{A}) - \frac{1}{c}\frac{d\vec{A}}{dt}\right)$$

The standard Lagrange equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_j}\right) - \frac{\partial L}{\partial x_j} = 0 \text{ where } L = T - V = \text{ Lagrangian}$$

works for a potential V which does not depend on the velocity.

However, when the force law takes the following form

$$F_j = -\frac{\partial U}{\partial x_j} + \frac{d}{dt}\left(\frac{\partial U}{\partial \dot{x}_j}\right)$$

we can still construct a Lagrange equation of the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_j}\right) - \frac{\partial L}{\partial x_j} = 0 \text{ where } L = T - U = \text{ Lagrangian}$$

We can get the Lorentz force law into this form by using a mathematical trick

$$\frac{dA_j}{dt} = \frac{d}{dt}\left(\frac{\partial}{\partial v_j}(\vec{v}\cdot\vec{A})\right) = \frac{d}{dt}\left(\frac{\partial}{\partial v_j}(\vec{v}\cdot\vec{A} - q\phi)\right)$$

where we have used the fact that the potential ϕ is independent of the velocity, to obtain

$$F_x = -\frac{\partial}{\partial x}\left(q\phi - \frac{q}{c}\vec{v}\cdot\vec{A}\right) + \frac{d}{dt}\left(\frac{\partial}{\partial v_x}\left(q\phi - \frac{q}{c}\vec{v}\cdot\vec{A}\right)\right)$$

This says that we need to choose the generalized potential U as

$$U = q\phi - \frac{q}{c}\vec{v}\cdot\vec{A}$$

and get the Lagrangian

$$L = T - U = \frac{1}{2}m\vec{v}^2 - q\phi + \frac{q}{c}\vec{v}\cdot\vec{A}$$

This gives the canonical momentum

$$\vec{p} = \frac{\partial L}{\partial \vec{v}} = m\vec{v} + \frac{q}{c}\vec{A}$$

The canonical momentum is NOT the simply the mechanical momentum but now includes an extra term proportional to the vector potential(the momentum of the field!).

The Hamiltonian is derived from the Lagrangian using

$$H = \vec{p} \cdot \vec{v} - L = \frac{1}{2m}\left(\vec{p} - \frac{q}{c}\vec{A}\right)^2 + q\varphi$$

This says that the simplest way of coupling the electromagnetic field to the motion of the particle is to replace the momentum by the quantity

$$\vec{p} - \frac{q}{c}\vec{A}$$

which includes the momentum associated with the electromagnetic field.

This method of introducing the electromagnetic field into the equations of motion of a particle is called minimal coupling.

We will see its relationship to gauge invariance shortly.

The transition to quantum mechanics is now done in the standard manner.

$$\hat{H}\,|\psi(t)\rangle = E\,|\psi(t)\rangle = ih\frac{\partial}{\partial t}\,|\psi(t)\rangle$$

$$\langle\vec{r}|\,\hat{H}\,|\psi(t)\rangle = E\,\langle\vec{r}\mid\psi(t)\rangle = ih\frac{\partial}{\partial t}\,\langle\vec{r}\mid\psi(t)\rangle$$

$$\langle\vec{r}|\,\hat{H}\,|\psi(t)\rangle = E\psi(\vec{r},t) = ih\frac{\partial\psi(\vec{r},t)}{\partial t}$$

where

$$\hat{H} = \frac{1}{2m}\left(\hat{\mathbf{p}} - \frac{q}{c}\vec{A}(\hat{\mathbf{r}},t)\right)^2 + q\phi(\hat{\mathbf{r}},t)$$

Using

$$\hat{\mathbf{p}} = \frac{\hbar}{i}\nabla$$

we then get the Schrodinger equation

$$\langle \vec{r}| \left[\frac{1}{2m}\left(\hat{\mathbf{p}} - \frac{q}{c}\vec{A}(\hat{\mathbf{r}},t)\right)^2 + q\phi(\hat{\mathbf{r}},t)\right] |\psi(t)\rangle = E\psi(\vec{r},t) = ih\frac{\partial\psi(\vec{r},t)}{\partial t}$$

$$\left[\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}(\vec{r},t)\right)^2 + q\phi(\vec{r},t)\right] \langle \vec{r} \mid \psi(t)\rangle = E\psi(\vec{r},t) = ih\frac{\partial\psi(\vec{r},t)}{\partial t}$$

$$\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}(\vec{r},t)\right)^2 \psi(\vec{r},t) + q\phi(\vec{r},t)\psi(\vec{r},t) = E\psi(\vec{r},t) = ih\frac{\partial\psi(\vec{r},t)}{\partial t}$$

The Schrodinger equation for a charged particle in an electromagnetic field must be invariant under a *gauge transformation* (since Maxwell's equations are so invariant), which is represented by the relations

$$\vec{A} \rightarrow \vec{A}' = \vec{A} + \nabla S(\vec{r},t)$$

$$\phi \rightarrow \phi' = \phi - \frac{1}{c}\frac{\partial S(\vec{r},t)}{\partial t}$$

where $S(\vec{r},t)$ is an arbitrary function.

In order for the Schrodinger equation to be invariant it must be transformed to

$$\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}'(\vec{r},t)\right)^2 \psi'(\vec{r},t) + q\phi'(\vec{r},t)\psi'(\vec{r},t) = ih\frac{\partial\psi'(\vec{r},t)}{\partial t}$$

where the wave function is only changed by an overall phase factor, i.e.,

$$\psi'(\vec{r},t) = \psi(\vec{r},t)e^{\frac{iq}{\hbar c}S(\vec{r},t)}$$

Proof:

$$\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}'(\vec{r},t)\right)^2 \psi'(\vec{r},t) + q\phi'(\vec{r},t)\psi'(\vec{r},t) = ih\frac{\partial\psi'(\vec{r},t)}{\partial t}$$

$$\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A} - \frac{q}{c}\nabla S\right)^2 \psi e^{\frac{iq}{\hbar c}S} + (q\phi - \frac{q}{c}\frac{\partial S}{\partial t})\psi e^{\frac{iq}{\hbar c}S} = ih\frac{\partial\psi e^{\frac{iq}{\hbar c}S}}{\partial t}$$

Using the identities

$$\frac{\partial\psi e^{\frac{iq}{\hbar c}S}}{\partial t} = \frac{\partial\psi}{\partial t}e^{\frac{iq}{\hbar c}S} + \frac{iq}{\hbar c}\frac{\partial S}{\partial t}\psi e^{\frac{iq}{\hbar c}S}$$

$$\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A} - \frac{q}{c}\nabla S\right)\left(\psi e^{\frac{iq}{\hbar c}S}\right) = e^{\frac{iq}{\hbar c}S}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}\right)\psi$$

we get the original Schrodinger equation back.

This result shows that the solutions of the gauge-transformed Schrodinger equation still describes the same physical states.

The wave functions or state vectors differ by a phase factor that depends on space and time and thus, the invariance is LOCAL rather than GLOBAL (a phase factor independent of space and time).

It is then clear that it is NOT the canonical momentum $\hat{\mathbf{p}} = -i\hbar\nabla$ (whose expectation value is *NOT gauge invariant*), but the genuine kinetic momentum

$$\hat{p} - \frac{q}{c}\vec{A}(\hat{r},t)$$

(whose expectation value IS gauge invariant), that represents a *measurable* quantity.

In any physical system, if the momentum operator $\hat{p}$ appears, then it must always be replaced by

$$\hat{p} - \frac{q}{c}\vec{A}(\hat{r}, t)$$

if we turn on an electromagnetic field.

This is the only way to guarantee gauge invariance in quantum mechanics.

Quantum mechanics + electromagnetism requires this *minimal coupling* for gauge invariance to be valid.

It is important to remember that local gauge transformations of the electromagnetic potentials requires that wave functions or state vectors transform via local phase changes, which means a phase factor that depends on space and time.

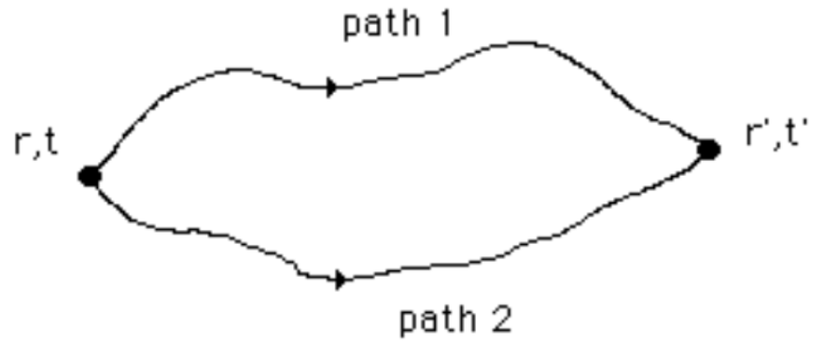
Feynman showed that the effect of turning on a magnetic field is to multiply the wave functions or probability amplitudes by the phase factor

$$\exp\left[\frac{iq}{\hbar c}\int_{path} d\vec{\ell}\cdot\vec{A}(\vec{r}, t)\right]$$

The Aharonov-Bohm Effect

The extra phase factor has a rather striking and unexpected consequence.

Let $\vec{A}$ be independent of *t* and consider the possible interference between the motions along two different paths as shown in the Figure.



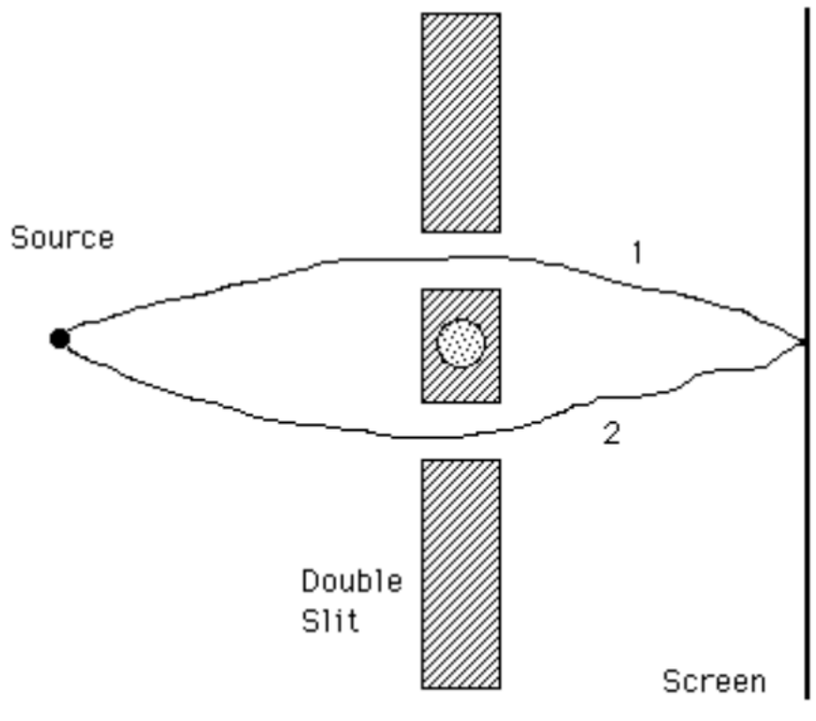
2-Path Experiment

When a magnetic field is turned on, the relative phase for the two paths (1 and 2) is changed by a factor

$$\frac{q}{\hbar c}\left[\int_{path1} d\vec{\ell}\cdot\vec{A} - \int_{path2} d\vec{\ell}\cdot\vec{A}\right] = \frac{q\Phi}{\hbar c}$$

where Φ = the total magnetic flux passing through the loop formed by the two paths or the flux enclosed by the loop (this is just Ampere's law).

Now consider the experimental setup shown in the next Figure.



Aharonov-Bohm Experiment

The cylindrical solenoid between the two slits generates a magnetic field out of the page and confined to the region of the solenoid.

This implies that the particles are traveling in regions *free* of any magnetic fields.

The *enclosed flux is NOT zero*, however.

As the flux is increased, the relative phase between the two paths changes as described above and the diffraction pattern on the screen changes.

This occurs even though there are no magnetic fields present along the paths.

There does exist, however, a nonzero vector potential in these regions and the relative phase is changing because the vector potential is changing.

This illustrates most powerfully that it is the electromagnetic potential, rather than the electromagnetic fields, as one might assume from Maxwell's equations, that are the fundamental physical quantities in quantum mechanics. This is called the *Aharonov-Bohm effect*.

Classically, the electric and magnetic fields, $\vec{\mathrm{E}}$ and $\vec{\mathrm{B}}$ are the physically relevant quantities, since they determine the Lorentz force.

In regions where the electric and magnetic fields are zero, a charged particle feels no force.

The vector and scalar potentials $\vec{\mathrm{A}}$ and ϕ serve in classical physics only as auxiliary quantities.

In quantum mechanics, the vector potential $\vec{\mathrm{A}}(\vec{\mathrm{r}})$ is the fundamental physical field.

The wave function, however, always has the property that physical variables and physical effects only depend on gauge-invariant quantities.

Time Evolution of Probabilities

Now that we have solved for the eigenvectors and eigenvalues of several systems we will return to a discussion of the time evolution of probability amplitudes.

Along the way we will reinforce the details about the most fundamental procedure in quantum mechanics.

Suppose we have some physical system that is describable by quantum mechanics.

We would like to be able to answer the following fundamental question?

> If the system is in the state $|\alpha\rangle$ at time $t = 0$,
> what is the probability that it will be in the
> state $|\beta\rangle$ at time t?

These two states are usually defined as follows:

$$|\alpha\rangle = \text{initial state} = \text{state prepared by experimenter}$$
$$|\beta\rangle = \text{final state} = \text{state defined by measuring device}$$

The formal answer to the question is

$$P_{\beta\alpha}(t) = |Amplitude_{\beta\alpha}(t)|^2 = |A_{\beta\alpha}(t)|^2 = \left|\langle\beta|\,\hat{U}(t)\,|\alpha\rangle\right|^2$$

where $\hat{U}(t)$ is the time evolution operator.

Let us assume for simplicity at this stage that the Hamiltonian $\hat{H}$ does not explicitly dependent on time t.

From our earlier discussions, we then have

$$\hat{U}(t) = e^{-i\hat{H}t/\hbar}$$

so that the probability amplitude becomes

$$A_{\beta\alpha}(t) = \langle\beta|\,\hat{U}(t)\,|\alpha\rangle = \langle\beta|\,e^{-\frac{i}{\hbar}\hat{H}t}\,|\alpha\rangle$$

The most important rule in quantum mechanics is the following:

> If you are evaluating an amplitude involving a particular operator, $\hat{H}$ in this case, then switch to a basis for the space corresponding to the eigenvectors of the that operator.

This basis is called the *HOME* space.

This is carried out by inserting identity operators as sums of *projection* operators written in the energy eigenvector basis:

$$\begin{aligned} A_{\beta\alpha}(t) &= \langle\beta|\,e^{-\frac{i}{\hbar}\hat{H}t}\,|\alpha\rangle = \langle\beta|\,\hat{I}e^{-\frac{i}{\hbar}\hat{H}t}\hat{I}\,|\alpha\rangle \\ &= \sum_{E}\sum_{E'}\langle\beta\mid E\rangle\langle E|\,e^{-\frac{i}{\hbar}\hat{H}t}\,|E'\rangle\langle E'\mid\alpha\rangle \end{aligned}$$

where

$$\begin{aligned} &\hat{H}\,|E\rangle = E\,|E\rangle \ \text{ and } \ \hat{H}\,|E'\rangle = E'\,|E'\rangle \\ &\langle E\mid E'\rangle = \delta_{EE'} \\ &\sum_{E}|E\rangle\langle E| = \hat{I} = \sum_{E'}|E'\rangle\langle E'| \end{aligned}$$

It is clear that the time evolution calculation is most easily done in this basis.

We then have

$$A_{\beta\alpha}(t) = \sum_E \sum_{E'} \langle \beta \mid E \rangle \langle E | e^{-\frac{i}{\hbar}E't} | E' \rangle \langle E' \mid \alpha \rangle$$

$$= \sum_E \sum_{E'} e^{-\frac{i}{\hbar}E't} \langle \beta \mid E \rangle \langle E \mid E' \rangle \langle E' \mid \alpha \rangle$$

$$= \sum_E \sum_{E'} e^{-\frac{i}{\hbar}E't} \langle \beta \mid E \rangle \delta_{EE'} \langle E' \mid \alpha \rangle$$

$$= \sum_E e^{-\frac{i}{\hbar}E't} \langle \beta \mid E \rangle \langle E \mid \alpha \rangle$$

Let us now describe a typical procedure:

1. Determine the initial state of the system $= |\alpha\rangle$
2. Expand the initial state in terms of the energy basis

$$|\alpha\rangle = \sum_{E''} |E''\rangle \langle E'' \mid \alpha \rangle$$

   which means we know the coefficients $\langle E'' \mid \alpha \rangle$
3. Expand the final state in terms of the energy basis

$$|\beta\rangle = \sum_{E''} |E''\rangle \langle E'' \mid \beta \rangle$$

   which means we know the coefficients $\langle E'' \mid \beta \rangle$
4. Calculate the amplitude and probability.

Example - Infinite Square Well System

Suppose that a particle in an infinite square well $[-a/2, +a/2]$ is initially in the state $|\alpha\rangle$, such that

$$\langle x \mid \alpha \rangle = \begin{cases} 0 & x < 0 \\ \frac{2}{\sqrt{a}} & 0 \le x \le \frac{a}{4} \\ 0 & x > \frac{a}{4} \end{cases}$$

Now we have found that

$$\langle x \mid E \rangle = \begin{cases} \frac{2}{\sqrt{a}} \cos \frac{n\pi x}{a} & n \text{ odd} \\ \frac{2}{\sqrt{a}} \sin \frac{n\pi x}{a} & n \text{ even} \end{cases}$$

where

$$E_n = \frac{n^2 \pi^2 \hbar^2}{2ma^2} \qquad n = 1, 2, 3, 4, \ldots$$

Therefore, we get

$$\langle E_n \mid \alpha \rangle = \int_{-\frac{a}{2}}^{\frac{a}{2}} \langle E_n \mid x \rangle \langle x \mid \alpha \rangle \, dx = \frac{2}{\sqrt{a}} \int_0^{\frac{a}{4}} \langle E_n \mid x \rangle \, dx$$

$$= \frac{2\sqrt{2}}{n\pi} \begin{cases} \sin \frac{n\pi}{4} & n \text{ odd} \\ -\cos \frac{n\pi}{4} & n \text{ even} \end{cases}$$

or

$$|\alpha\rangle = \langle E_1 \mid \alpha \rangle |E_1\rangle + \langle E_2 \mid \alpha \rangle |E_2\rangle + \langle E_3 \mid \alpha \rangle |E_3\rangle + \ldots$$

$$= \frac{2}{\pi} |E_1\rangle + \frac{2}{3\pi} |E_3\rangle + \frac{2\sqrt{2}}{4\pi} |E_4\rangle + \frac{2}{5\pi} |E_5\rangle - \frac{2}{7\pi} |E_7\rangle - \frac{2\sqrt{2}}{8\pi} |E_8\rangle + \ldots$$

If we assume that the final state $|\beta\rangle$ is an energy eigenstate, say $|E_n\rangle$, we then have the probability amplitude that the particle started out in state $|\alpha\rangle$ at $t = 0$ is found in the state $|E_n\rangle$ at time t is

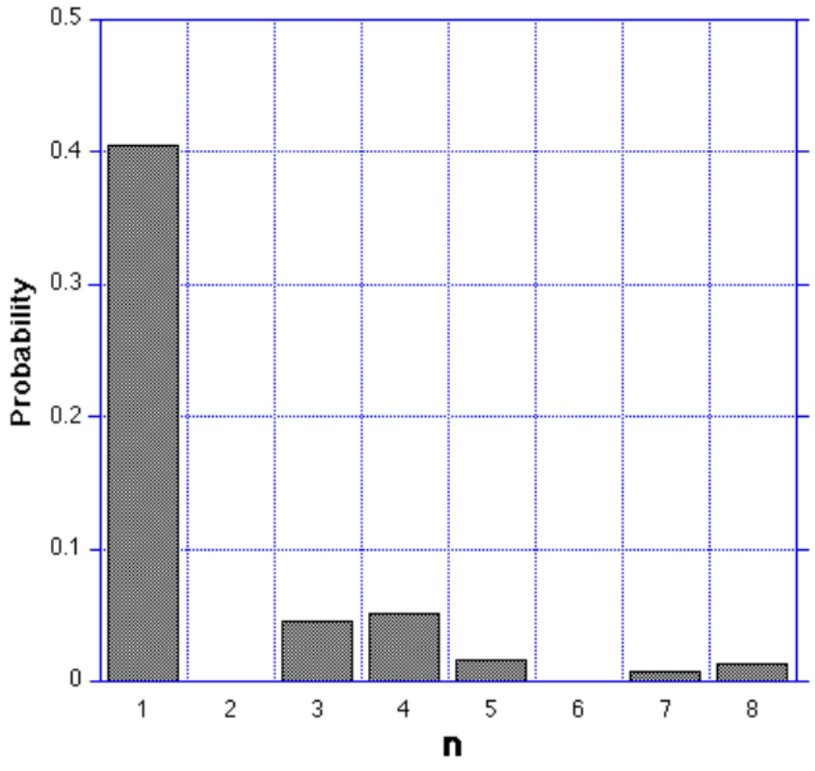
$$A_{n\alpha}(t) = \sum_E e^{-\frac{i}{\hbar} Et} \langle E_n \mid E \rangle \langle E \mid \alpha \rangle = e^{-\frac{i}{\hbar} E_n t} \langle E_n \mid \alpha \rangle$$

and the corresponding probabilities are

$$P_{n\alpha}(t) = |\langle E_n \mid \alpha \rangle|^2$$

which is independent of time.

A plot of these fixed probabilities versus n is shown in the Figure below.



Probabilities

If, on the other hand, we assume that the final state is a linear combination of two energy eigenstates, say,

$$|\beta\rangle = \frac{1}{\sqrt{2}}|E_1\rangle + \frac{1}{\sqrt{2}}|E_3\rangle$$

then we have the probability amplitude that the particle started out in state $|\alpha\rangle$ at t=0 is found in the state $|\beta\rangle$ at time t is

$$\begin{aligned} A_{\beta\alpha}(t) &= \sum_E e^{-\frac{i}{\hbar}Et}\langle\beta \mid E\rangle\langle E \mid \alpha\rangle \\ &= \frac{1}{\sqrt{2}}\left(e^{-\frac{i}{\hbar}E_1 t}\langle E_1 \mid \alpha\rangle + e^{-\frac{i}{\hbar}E_3 t}\langle E_3 \mid \alpha\rangle\right) \\ &= \frac{1}{\sqrt{2}}|E_1\rangle + \frac{1}{\sqrt{2}}|E_3\rangle \end{aligned}$$

and the corresponding probability is

$$P_{\beta\alpha}(t) = |\langle\beta \mid \alpha\rangle|^2 = \frac{1}{2}\left|e^{-\frac{i}{\hbar}E_1 t}\langle E_1 \mid \alpha\rangle + e^{-\frac{i}{\hbar}E_3 t}\langle E_3 \mid \alpha\rangle\right|^2$$

$$= \frac{1}{2}\left|e^{-\frac{i}{\hbar}E_1 t}\frac{2}{\pi} + e^{-\frac{i}{\hbar}E_3 t}\frac{2}{3\pi}\right|^2 = \frac{2}{\pi^2}\left|\frac{1}{3} + e^{\frac{i}{\hbar}(E_3 - E_1)t}\right|^2$$

$$= \frac{2}{\pi^2}\left(\frac{1}{9} + \frac{2}{3}\cos\left[\frac{1}{\hbar}(E_3 - E_1)t\right] + 1\right)$$

$$= \frac{20}{9\pi^2}\left(1 + \frac{3}{5}\cos\left[\frac{1}{\hbar}(E_3 - E_1)t\right]\right)$$

which is dependent on time.

In this case the probability of observing this final state oscillates between 0.18 and 0.72.

Translation Invariant Potential - Bands

We now consider a potential that is *periodic in space*, which means that it is *translationally invariant*, i.e.,

$$V(x) = V(x + a) \text{ where } a = \text{ some fixed distance}$$

This might represent an electron in a 1-dimensional crystal lattice, where a is the distance between atoms.

The change in the allowed energy spectrum caused by the periodicity of the lattice is rather striking.

The Hamiltonian of the system is

$$\hat{H}(x, p) = \frac{p^2}{2m} + V(x)$$

This Hamiltonian is invariant if we translate a fixed distance a

$$\hat{H}(x + a, p) = \frac{p^2}{2m} + V(x + a) = \frac{p^2}{2m} + V(x) = \hat{H}(x, p)$$

This means that the transformation operator $\hat{T}(a)$ that corresponds to a translation by a distance a *commutes* with the Hamiltonian

$$\left[\hat{H}, \hat{T}(a)\right] = 0$$

and thus they must *share a set of eigenfunctions*.

From our earlier discussion, the translation operator has the defining property that, if the state $|\psi\rangle$ has the wave function $\langle x|\psi\rangle$ in the position representation, then the state $\hat{T}(a)|\psi\rangle$ has the wave function

$$\langle x|\hat{T}(a)|\psi\rangle = \langle x + a \mid \psi\rangle$$

Let us rederive the eigenfunctions of $\hat{T}(a)$.

If $|\psi\rangle$ is an eigenstate of the $\hat{T}(a)$ operator, then we must have

$$\hat{T}(a)|\psi\rangle = \lambda|\psi\rangle \text{ where } \lambda = \text{ the eigenvalue}$$

and thus

$$\psi(x + a) = \langle x|\hat{T}(a)|\psi\rangle = \lambda\langle x \mid \psi\rangle = \lambda\psi(x)$$

With hindsight, we write λ in the form e^{ika} which defines k.

We then have

$$\psi(x + a) = e^{ika}\psi(x)$$

If we define

$$u_k(x) = e^{-ikx}\psi(x)$$

we then have that

$$\psi(x + a) = e^{ik(x+a)}u_k(x + a) = e^{ika}\psi(x) = e^{ika}e^{ikx}u_k(x)$$
$$u_k(x + a) = u_k(x)$$

which says that $u_k(x)$ is also periodic in space with period a.

The eigenfunctions of $\hat{T}$ (a) are then given by

$$\psi(x) = e^{ikx} u_k(x)$$

which is a *plane wave modulated by a function with the periodicity of the lattice* (potential).

This type of wave function is called a *Bloch wave*.

We note that k must be real since if k were not real, then the eigenfunction would diverge as x gets large and thus be unrenormalizable, which is not allowed.

Since $\hat{H}$ and $\hat{T}$ (a) must share a set of eigenfunctions, we can construct a complete set of energy eigenstates in the form of Bloch waves.

This means that we can find at least one energy eigenstate for each possible $\lambda = e^{ika}$, since all λ are eigenvalues of $\hat{T}$ (a).

Note also that the values k and $k + 2\pi\, n/a$ give the same eigenvalue and thus we can restrict our attention to k values in the interval

$$-\frac{\pi}{a} \leq k \leq \frac{\pi}{a}$$

and we will be including all possible eigenvalues of $\hat{T}$ (a).

Example

Let us now consider the specific example of a 1-dimensional periodic array of delta functions

$$V(x) = \sum_{n=-\infty}^{\infty} v_0 \delta(x - na)$$

which corresponds to the so-called *Kronig-Penney model*.

In the interval $0 < x < a$ the potential vanishes and the solution to the Schrodinger equation is of the form

$$\psi(x) = Ae^{iqx} + Be^{-iqx} \text{ where } E = \frac{\hbar^2 q^2}{2m}$$

Therefore, in the interval $0 < x < a$, we must have

$$u_k(x) = Ae^{i(q-k)x} + Be^{-i(q+k)x}$$

The coefficients A and B are fully determined by two conditions:

1. $\psi(x)$ and hence $u_k(x)$ must be continuous at the lattices sites

$$\lim_{\eta \to 0} (u_k(\eta) - u_k(-\eta)) \to 0$$

Periodicity then says that $u_k(-\eta) = u_k(a-\eta)$ and hence we must also have

$$\lim_{\eta \to 0} (u_k(\eta) - u_k(a-\eta)) \to 0$$

which gives

$$A + B = Ae^{i(q-k)a} + Be^{-i(q+k)a} \qquad \text{(condition 1)}$$

2. $d\psi(x)/dx$ is discontinuous at the lattice points, as we showed earlier, as follows:

$$\lim_{\eta \to 0} \frac{\hbar^2}{2m} \left(\left. \frac{d\psi(x)}{dx} \right|_{x=\eta} - \left. \frac{d\psi(x)}{dx} \right|_{x=-\eta} \right) = v_0 \psi(0)$$

We have

$$\lim_{\eta\to 0}\left(\left.\frac{d\psi(x)}{dx}\right|_{x=\eta}\right) = iq(A-B)$$

$$\lim_{\eta\to 0}\left(\left.\frac{d\psi(x)}{dx}\right|_{x=-\eta}\right) = e^{-ika}\lim_{\eta\to 0}\left(\left.\frac{d\psi(x)}{dx}\right|_{x=a-\eta}\right)$$

$$= e^{-ika}iq\left(Ae^{iqa} - Be^{-iqa}\right)$$

$$\psi(0) = A + B$$

which gives

$$\frac{\hbar^2}{2m}iq\left(A - B - Ae^{i(q-k)a} + Be^{-i(q+k)a}\right)$$

$$= v_0(A+B) \qquad \text{(condition 2)}$$

We only need these two conditions and do not need to worry about all of the rest of the potential because of the periodicity.

We solve for the ratio A/B in each equation and equate the two results to get a transcendental equation for q in terms of k (which then gives us the allowed energy values).

We get the equation

$$\cos ka = \cos qa + \frac{mv_0}{q\hbar^2}\sin qa$$

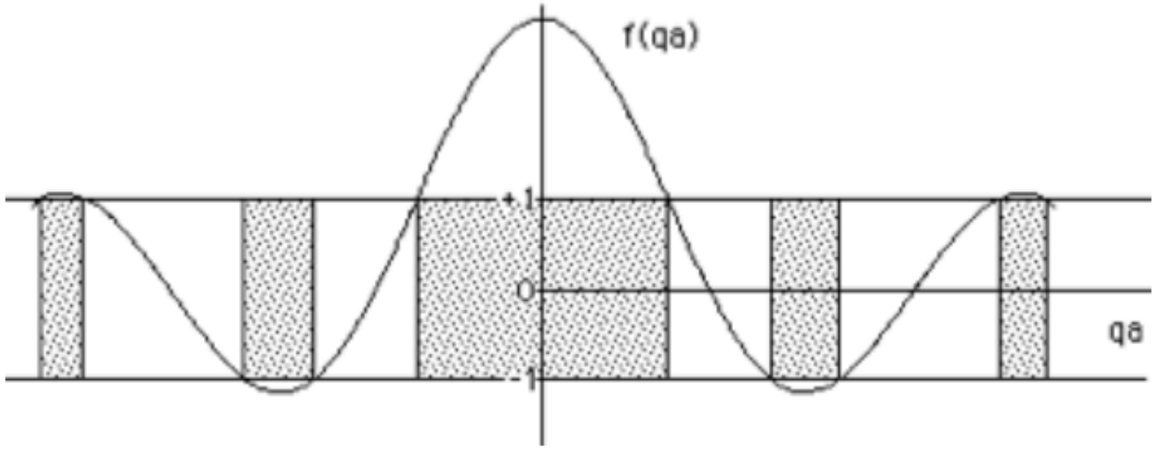
Now for any given value of k we must have

$$-1 \le \cos ka \le +1$$

Therefore, if we plot

$$\cos ka = \cos qa + \frac{mv_0 a}{\hbar^2}\frac{\sin qa}{qa} \quad \text{versus} \quad qa$$

the only allowed q values are between the two lines representing ±1 as shown in the Figure below.



Allowed q values

i.e., q must always lie in the unshaded areas.

For each value of cos ka there are an infinite number of solutions.

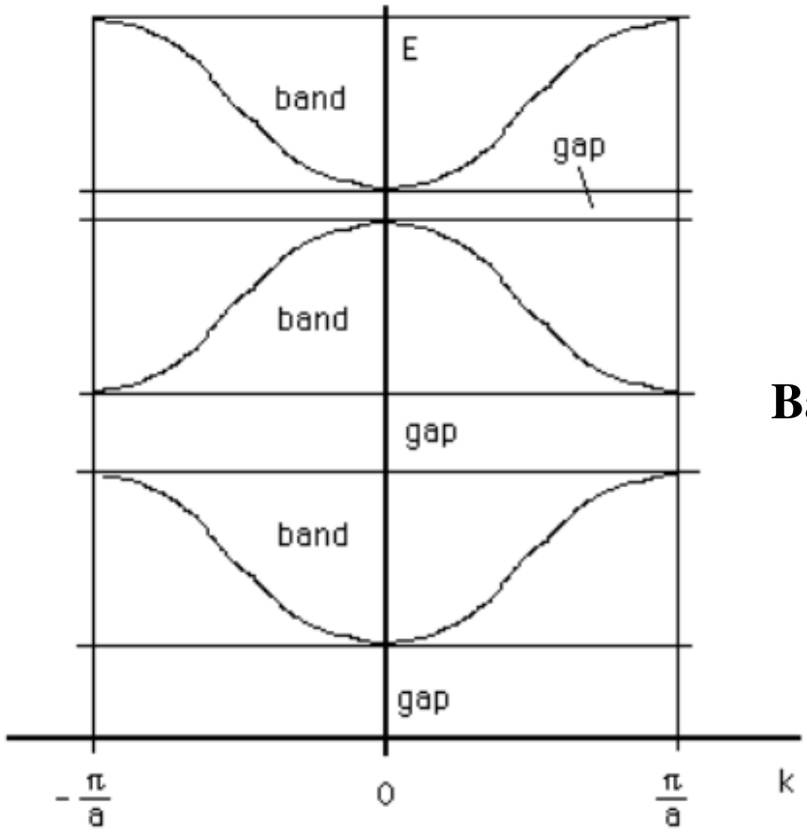
Taking k in the range $-\pi/a \le k \le \pi/a$, we can plot

$$E = \frac{\hbar^2 q^2}{2m} \quad \text{versus} \quad ka$$

as shown in the Figure.

The possible energy values lie in bands, with gaps between them.

This structure of the energy spectrum with bands and gaps occurs for all periodic potentials, which is s dominant feature in Solid State Physics.



Bands and Gaps

Chapter 7 - Angular Momentum; 2- and 3-Dimensions

Angular Momentum Eigenvalues and Eigenvectors

Earlier, we derived the commutation relations that define angular momentum operators

$$\left[\hat{J}_i, \hat{J}_j\right] = i\hbar\varepsilon_{ijk}\hat{J}_k \equiv i\hbar\sum_k \varepsilon_{ijk}\hat{J}_k$$

where

$$\epsilon_{ijk} = \begin{cases} +1 & \text{if } ijk = \text{even permutation of } 123 \\ -1 & \text{if } ijk = \text{odd permutation of } 123 \\ 0 & \text{if any two indices are identical} \end{cases}$$

and the *Einstein summation convention over repeated indices* is understood if the summation sign is left out (unless an explicit override is given).

In addition, these are all Hermitian operators, i.e., $(\hat{J}_i)^\dagger = \hat{J}_i^\dagger = \hat{J}_i$.

Since these three operators form a *closed commutator algebra* (no other operators are involved), we can solve for the eigenvectors and eigenvalues using only the commutators.

The three operators $\hat{J}_1$, $\hat{J}_2$ and $\hat{J}_3$ do not commute with each other and hence do not share a common set of eigenvectors (called a *representation*).

However, there exists another operator that commutes with each of the angular momentum components separately.

If we define

$$\hat{J}^2 = \sum_{i=1}^{3} \hat{J}_i^2$$

which is the square of the total angular momentum vector, we have

$$\left[\hat{J}^2, \hat{J}_i\right] = 0 \text{ for } i = 1, 2, 3$$

In addition, we have $(\hat{J}^2)^\dagger = \hat{J}^2$, so that $\hat{J}^2$ is Hermitian also.

The commutation relations and the Hermitian property say that $\hat{J}^2$ and any one of the components *share a complete set of common eigenvectors*.

By convention, we choose to use $\hat{J}^2$ and $\hat{J}_3$ as the two operators, whose eigenvalues(good quantum numbers) will characterize the set of common eigenvectors.

As we shall see, $\hat{J}^2$ is a so-called *Casimir invariant operator* that characterizes the representations(set of eigenvectors).

In particular, the eigenvalue of $\hat{J}^2$ characterizes the representation and the eigenvalues of one of the components of the angular momentum (usually $\hat{J}_3$) will *characterize the eigenvectors within a representation*.

We define the eigenvector/eigenvalue relations by the equations

$$\hat{J}^2 |\lambda m\rangle = \lambda \hbar^2 |\lambda m\rangle$$
$$\hat{J}_3 |\lambda m\rangle = m\hbar |\lambda m\rangle$$

where the appropriate factors of $\hbar$ that have been explicitly put into the relations make m and λ dimensionless numbers.

We now define some other operators and their associated commutators so that we can use them in our derivations.

$$\hat{J}_\pm = \hat{J}_1 \pm i\hat{J}_2$$
$$\hat{J}_- = \left(\hat{J}_+\right)^+ \rightarrow \text{ they are not Hermitian operators}$$

We then have

$$\left[\hat{J}^2, \hat{J}_\pm\right] = \left[\hat{J}^2, \hat{J}_1\right] \pm i\left[\hat{J}^2, \hat{J}_2\right] = 0$$

$$\left[\hat{J}_3, \hat{J}_\pm\right] = \left[\hat{J}_3, \hat{J}_1\right] \pm i\left[\hat{J}_3, \hat{J}_2\right]$$

$$= i\hbar\hat{J}_2 \mp i\left(i\hbar\hat{J}_1\right) = \pm\hbar\hat{J}_\pm$$

and

$$\left[\hat{J}_+, \hat{J}_-\right] = \left[\hat{J}_1, \hat{J}_1\right] + i\left[\hat{J}_2, \hat{J}_1\right] - i\left[\hat{J}_1, \hat{J}_2\right] - \left[\hat{J}_2, \hat{J}_2\right]$$

$$= -2i\left[\hat{J}_1, \hat{J}_2\right] = 2\hbar\hat{J}_3$$

and

$$\hat{J}_+\hat{J}_- = \left(\hat{J}_1 + i\hat{J}_2\right)\left(\hat{J}_1 - i\hat{J}_2\right)$$

$$= \hat{J}_1^2 + \hat{J}_2^2 - i\left[\hat{J}_1, \hat{J}_2\right] = \hat{J}^2 - \hat{J}_3^2 + \hbar\hat{J}_3$$

and

$$\hat{J}_-\hat{J}_+ = \hat{J}^2 - \hat{J}_3^2 - \hbar\hat{J}_3$$

Finally, we have

$$\hat{J}^2 = \frac{\hat{J}_+\hat{J}_- + \hat{J}_-\hat{J}_+}{2} + \hat{J}_3^2$$

Derivation of Eigenvalues

Now the definitions above tell us that

$$\langle \lambda m | \hat{J}^2 | \lambda m \rangle = \lambda\hbar^2 \langle \lambda m \mid \lambda m \rangle = \sum_i \langle \lambda m | \hat{J}_i^2 | \lambda m \rangle$$

$$\lambda\hbar^2 \langle \lambda m \mid \lambda m \rangle = \sum_i \langle \lambda m | \hat{J}_i\hat{J}_i | \lambda m \rangle = \sum_i \langle \lambda m | \hat{J}_i^+\hat{J}_i | \lambda m \rangle$$

Let us define the new vector $|\alpha_i\rangle = \hat{J}_i |\lambda m\rangle$.

Remember that the norm of any vector is non-negative, i.e., $\langle a \mid a\rangle \geq 0$.

Therefore

$$\langle \lambda m \mid \lambda m\rangle \geq 0 \text{ and } \langle \alpha_i \mid \alpha_i\rangle \geq 0$$

Now since $\langle\alpha_i| = \langle\lambda m| \hat{J}^\dagger_i$ we have

$$\lambda\hbar^2 \langle \lambda m \mid \lambda m\rangle = \sum_i \langle \lambda m| \hat{J}_i^+ \hat{J}_i |\lambda m\rangle = \sum_i \langle \alpha_i \mid \alpha_i\rangle \geq 0$$

or we have

$$\lambda \geq 0 \text{ or the eigenvalues of } \hat{J}^2 \text{ are greater than } 0$$

This says that for a fixed value of λ (the eigenvalue of $\hat{J}^2$), which characterizes the representation, there must be maximum and minimum values of m (the eigenvalue of $\hat{J}_3$), which characterizes the eigenvectors within a representation.

Now we have

$$\begin{aligned}\hat{J}_3\hat{J}_+ |\lambda m\rangle &= \hat{J}_3(\hat{J}_+ |\lambda m\rangle) \\ &= (\hat{J}_+\hat{J}_3 + [\hat{J}_3, \hat{J}_+]) |\lambda m\rangle = (\hat{J}_+\hat{J}_3 + \hbar\hat{J}_+) |\lambda m\rangle \\ &= \hbar(m+1)\hat{J}_+ |\lambda m\rangle = \hbar(m+1)(\hat{J}_+ |\lambda m\rangle)\end{aligned}$$

which says that $\hat{J}_+ |\lambda m\rangle$ is an eigenvector of $\hat{J}_3$ with the *raised* eigenvalue $\hbar$ (m+1), i.e., $\hat{J}_+ |\lambda m\rangle \propto |\lambda, m+1\rangle$ (remember the harmonic oscillator discussion).

Since we already showed that for fixed λ, there must be a maximum value of m, say $m_{\max}$, then it must be the case that for that particular m-value we have

$$\hat{J}_+ |\lambda m_{\max}\rangle = 0$$

If this were not true, then we would have

$$\hat{J}_+ |\lambda m_{\max}\rangle \propto |\lambda, m_{\max}+1\rangle$$

but this violates the statement that $m_{\max}$ was was the maximum m-value.

Using this result we find

$$\hat{J}_-\hat{J}_+ |\lambda m_{\max}\rangle = 0 = \left(\hat{J}^2 - \hat{J}_3^2 - \hbar\hat{J}_3\right)|\lambda m_{\max}\rangle$$
$$\hbar^2\left(\lambda - m_{\max}^2 - m_{\max}\right)|\lambda m_{\max}\rangle = 0$$
$$\lambda - m_{\max}^2 - m_{\max} = 0 \rightarrow \lambda = m_{\max}^2 + m_{\max}$$
$$\lambda = m_{\max}(m_{\max}+1)$$

It is convention to define

$$m_{\max} = j \text{ and hence } \lambda = j(j+1)$$

In the same way we can show

$$\hat{J}_3\hat{J}_- |\lambda m\rangle = \hat{J}_3(\hat{J}_- |\lambda m\rangle)$$
$$= (\hat{J}_-\hat{J}_3 + [\hat{J}_3, \hat{J}_-])|\lambda m\rangle = (\hat{J}_-\hat{J}_3 - \hbar\hat{J}_+)|\lambda m\rangle$$
$$= \hbar(m-1)\hat{J}_- |\lambda m\rangle = \hbar(m-1)(\hat{J}_- |\lambda m\rangle)$$

which says that $\hat{J}_-|\lambda m\rangle$ is an eigenvector of $\hat{J}_3$ with the *lowered* eigenvalue $\hbar(m-1)$, i.e., $\hat{J}_-|\lambda m\rangle \propto |\lambda, m-1\rangle$.

If we let the minimum value of m be $m_{\min}$, then as before we must have

$$\hat{J}_- |\lambda m_{\min}\rangle = 0$$

or $m_{\min}$ is not the minimum value of m.

This says that

$$\hat{J}_+\hat{J}_-\left|\lambda m_{\min}\right\rangle = 0 = \left(\hat{J}^2 - \hat{J}_3^2 + \hbar\hat{J}_3\right)\left|\lambda m_{\min}\right\rangle$$

$$\hbar^2\left(\lambda - m_{\min}^2 + m_{\min}\right)\left|\lambda m_{\min}\right\rangle = 0$$

$$\lambda - m_{\min}^2 + m_{\min} = 0 \rightarrow \lambda = m_{\min}^2 - m_{\min}$$

$$\lambda = m_{\min}(m_{\min} - 1) = j(j+1)$$

which says that

$$m_{\min} = -j$$

We have thus shown that the pair of operators $\hat{J}^2$ and $\hat{J}_3$ have a common set of eigenvectors $|jm\rangle$ (we now use the labels j and m), where we have found that

$$-j \le m \le j$$

and the allowed m-values change by steps of one, i.e., for a given j-value, the allowed m-values are

$$-j, -j+1, -j+2, \ldots\ldots, j-2, j-1, j$$

which implies that

$$2j = \text{ integer}$$

or

$$j = \frac{\text{integer}}{2} \ge 0 \quad \text{the allowed values}$$

Thus, we have the *allowed sets or representations of the angular momentum commutation relations* given by

$$j = 0 \quad , \quad m = 0$$

$$j = \frac{1}{2} \quad , \quad m = \frac{1}{2}, -\frac{1}{2}$$

$$j = 1 \quad , \quad m = 1, 0, -1$$

$$j = \frac{3}{2} \quad , \quad m = \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$$

and so on.

For each value of j, there are $2j + 1$ allowed m-values in the eigenvalue spectrum(representation) and

$$\hat{J}^2 |jm\rangle = \hbar^2 j(j+1) |jm\rangle$$

$$\hat{J}_3 |jm\rangle = m\hbar |jm\rangle$$

Before proceeding, we need a few more relations.

We found earlier that

$$\hat{J}_+ |jm\rangle = C_+ |j, m+1\rangle = |\alpha_+\rangle$$

$$\hat{J}_- |jm\rangle = C_- |j, m-1\rangle = |\alpha_-\rangle$$

and from these we have

$$\langle\alpha_+| = C_+^* \langle j, m+1|$$

$$\langle\alpha_-| = C_-^* \langle j, m-1|$$

We can then say that

$$
\begin{aligned}
\langle \alpha_+ \mid \alpha_+ \rangle &= |C_+|^2 \langle j, m+1 \mid j, m+1 \rangle = |C_+|^2 \\
&= (\langle jm| (\hat{J}_+)^+)(\hat{J}_+ |jm\rangle) \\
&= \langle jm| \hat{J}_- \hat{J}_+ |jm\rangle \\
&= \langle jm| \left(\hat{J}^2 - \hat{J}_3^2 - \hbar \hat{J}_3\right) |jm\rangle \\
&= \langle jm| \hbar^2 (j(j+1) - m^2 - m) |jm\rangle \\
&= \hbar^2 (j(j+1) - m^2 - m) \langle jm \mid jm \rangle \\
&= \hbar^2 (j(j+1) - m^2 - m)
\end{aligned}
$$

or

$$
C_+ = \hbar \sqrt{j(j+1) - m(m+1)}
$$

and similarly

$$
\begin{aligned}
\langle \alpha_- \mid \alpha_- \rangle &= |C_-|^2 \langle j, m-1 \mid j, m-1 \rangle = |C_-|^2 \\
&= (\langle jm| (\hat{J}_-)^+)(\hat{J}_- |jm\rangle) \\
&= \langle jm| \hat{J}_+ \hat{J}_- |jm\rangle \\
&= \langle jm| \left(\hat{J}^2 - \hat{J}_3^2 + \hbar \hat{J}_3\right) |jm\rangle \\
&= \langle jm| \hbar^2 (j(j+1) - m^2 + m) |jm\rangle \\
&= \hbar^2 (j(j+1) - m^2 + m) \langle jm \mid jm \rangle \\
&= \hbar^2 (j(j+1) - m^2 + m)
\end{aligned}
$$

or

$$C_- = \hbar\sqrt{j(j+1) - m(m-1)}$$

Therefore, we have the very important relations for the raising/lowering or ladder operators

$$\hat{J}_\pm |jm\rangle = \hbar\sqrt{j(j+1) - m(m\pm 1)}\,|j, m\pm 1\rangle$$
$$= \hbar\sqrt{(j\pm m+1)(j\mp m)}\,|j, m\pm 1\rangle$$

Transformations and Generators; Spherical Harmonics

There are many vector operators and vector component operators in the following discussions.

To avoid confusing notation, we will adopt the following conventions:

$\vec{A}_{op}$ = a vector operator

$\hat{A}_j$ = a vector component operator

$\hat{B}$ = any non - vector operator or it might be a unit vector (context will decide)

$\vec{q}$ = an ordinary vector

As we showed earlier, the angular momentum operators are the generators of rotations.

The unitary transformation operator for a rotation through an angle θ about an axis along the direction specified by the unit vector $\hat{n}$ is given by

$$\hat{U}_{\hat{n}}(\theta) = e^{-\frac{i}{h}\theta\hat{n}\cdot\vec{J}_{op}}$$

where $\vec{J}_{op}$ is the angular momentum operator.

What does the $\hat{J}_3$ operator look like in the position representation?

We will quickly get an idea of the answer and then step back and do things in more detail.

Suppose that we have a position representation wave function $\psi(\vec{r})$.

For a rotation about the 3-axis, we showed earlier that for an infinitesimal angle ε

$$\hat{U}_3(\varepsilon)\psi(x_1, x_2, x_3) = \psi(x_1 \cos\varepsilon + x_2 \sin\varepsilon, -x_1 \sin\varepsilon + x_2 \cos\varepsilon, x_3)$$

In general, for infinitesimal shifts(a in the case below) in the coordinates, we have, to lowest order in the infinitesimals, for a function of one variable

$$f(x + a) = f(x) + a\frac{\partial f}{\partial x}$$

Extending this to two variables, we have, to first order in infinitesimals a and b,

$$f(x + a, y + b) = f(x, y) + a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial y}$$

Therefore, for an infinitesimal angle of rotation ε,

$$\begin{aligned}\hat{U}_3(\varepsilon)\psi(x_1, x_2, x_3) &= \psi(x_1 \cos\varepsilon + x_2 \sin\varepsilon, -x_1 \sin\varepsilon + x_2 \cos\varepsilon, x_3) \\ &= \psi(x_1, x_2, x_3) + \varepsilon x_2 \frac{\partial\psi}{\partial x_1} - \varepsilon x_1 \frac{\partial\psi}{\partial x_2} \\ &= \left(1 + \varepsilon\left(x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2}\right)\right)\psi(x_1, x_2, x_3)\end{aligned}$$

But we also have (to first order)

$$\hat{U}_3(\varepsilon)\psi(x_1, x_2, x_3) = (1 - \frac{i}{\hbar}\varepsilon\hat{J}_3)\psi(x_1, x_2, x_3)$$

Putting these two equations together we have

$$\hat{J}_3 = -i\hbar\left(x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2}\right) = (\vec{r}_{op}\times(-i\hbar\nabla))_3$$
$$= (\vec{r}_{op}\times\vec{p}_{op})_3 = \left(\vec{L}_{op}\right)_3 = \hat{L}_3$$

where

$\vec{L}_{op}$ = orbital angular momentum operator

$\vec{r}_{op} = (\hat{x},\hat{y},\hat{z}) = (\hat{x}_1,\hat{x}_2,\hat{x}_3)$

$\vec{p}_{op} = (\hat{p}_x,\hat{p}_y,\hat{p}_z) = (\hat{p}_1,\hat{p}_2,\hat{p}_3)$

Since $\vec{L}_{\text{op}}$ is an angular momentum, it must have the commutation relations

$$\left[\hat{L}_i,\hat{L}_j\right] = i\hbar\varepsilon_{ijk}\hat{L}_k$$

where as we indicated above

$$\hat{L}_1 = \hat{x}_2\hat{p}_3 - \hat{x}_3\hat{p}_2$$
$$\hat{L}_2 = \hat{x}_3\hat{p}_1 - \hat{x}_1\hat{p}_3$$
$$\hat{L}_3 = \hat{x}_1\hat{p}_2 - \hat{x}_2\hat{p}_1$$

and

$$\vec{L}_{op}^{+} = \vec{L}_{op}$$

Using

$$\left[\hat{x}_i,\hat{p}_j\right] = i\hbar\delta_{ij}$$

we get

$$\left[\hat{L}_i,x_j\right] = i\hbar\varepsilon_{ijk}x_k$$

and for $\hat{n}$ = a unit vector

$$\sum_i \hat{n}_i \left[\hat{L}_i, x_j\right] = i\hbar \sum_i \hat{n}_i \varepsilon_{ijk} x_k$$

$$\left[\hat{n}\cdot\vec{L}_{op}, x_j\right] = i\hbar \left(\vec{r}_{op}\times\hat{n}\right)_j$$

$$\left[\hat{n}\cdot\vec{L}_{op}, \vec{r}_{op}\right] = \sum_j \left[\hat{n}\cdot\vec{L}_{op}, x_j\right]\hat{e}_j$$

$$= i\hbar \sum_j \left(\vec{r}_{op}\times\hat{n}\right)_j \hat{e}_j = i\hbar \left(\vec{r}_{op}\times\hat{n}\right)$$

where we have used

$$\vec{A}\times\vec{B} = \sum_{ijk} \varepsilon_{ijk} A_j B_k \hat{e}_i$$

Similarly, we get

$$\left[\hat{n}\cdot\vec{L}_{op}, \vec{p}_{op}\right] = i\hbar \left(\vec{p}_{op}\times\hat{n}\right)$$

Now let us step back and consider rotations in 3-dimensional space and try to get a better physical understanding of what is happening.

Consider the operator

$$\vec{r}\,'_{op} = \vec{r}_{op} + \vec{\alpha}\times\vec{r}_{op}$$

where $\vec{\alpha}$ = ordinary infinitesimal vector.

Now let $|\vec{r}_0\rangle$ = an eigenstate of $\vec{r}_{op}$ with eigenvalue $\vec{r}_0$, i.e.,

$$\vec{r}_{op}\,|\vec{r}_0\rangle = \vec{r}_0\,|\vec{r}_0\rangle$$

We then have

$$\vec{r}\,'_{op}\,|\vec{r}_0\rangle = \vec{r}\,'_0\,|\vec{r}_0\rangle = \left(\vec{r}_0 + \vec{\alpha}\times\vec{r}_0\right)|\vec{r}_0\rangle$$

which says that $|\vec{r}_0\rangle$ is also an eigenstate of $\vec{r}\,'_{op}$ (different eigenvalue).

Now, for simplicity, let $\vec{\alpha} = \alpha\,\hat{e}_3$, which says that

$$\vec{\alpha} \times \vec{r}_0 = \alpha\hat{e}_3 \times (x_{01}\hat{e}_1 + x_{02}\hat{e}_2 + x_{03}\hat{e}_3) = \alpha x_{01}\hat{e}_2 - \alpha x_{02}\hat{e}_1$$

$$\vec{r}_0 + \vec{\alpha} \times \vec{r}_0 = (x_{01} - \alpha x_{02})\hat{e}_1 + (x_{02} + \alpha x_{01})\hat{e}_2 + x_{03}\hat{e}_3$$

This last expression is the vector we get if we rotate $\vec{r}_0$ about $\hat{\alpha} = \vec{\alpha}/|\vec{\alpha}|$ ($\hat{e}_3$ in this case) by an infinitesimal angle $\hat{\alpha} = |\vec{\alpha}|$.

This result generalizes for $\vec{\alpha}$ in any direction.

Since the eigenvalues of $\vec{r}'_{op}$ are those of $\vec{r}_{op}$ rotated by $|\vec{\alpha}|$ about $\vec{\alpha}$, we conclude that $\vec{r}'_{op}$ is the position operator rotated by rotated by $|\vec{\alpha}|$ about $\vec{\alpha}$.

Alternatively, $\vec{r}'_{op}$ is the position operator in a coordinate frame rotated by $|\vec{\alpha}|$ about $\vec{\alpha}$ (corresponds to active/passive views).

To connect this to generators, unitary transformations, and angular momentum, we proceed as follows. We can rewrite $\vec{r}'_{op}$ as

$$\vec{r}\,'_{op} = \vec{r}_{op} + \frac{i}{\hbar}\left[\vec{\alpha}\cdot\vec{L}_{op}, \vec{r}_{op}\right]$$

which is equivalent (to first order in α) to

$$\vec{r}\,'_{op} = e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\vec{r}_{op}e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} \qquad \text{i.e.,} \quad \vec{r}\,'_{op} = e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\vec{r}_{op}e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}$$

$$= (\hat{I} + \frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op})\vec{r}_{op}(\hat{I} - \frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}) + O(\alpha^2)$$

$$= \vec{r}_{op} + \frac{i}{\hbar}\left[\vec{\alpha}\cdot\vec{L}_{op}, \vec{r}_{op}\right] + O(\alpha^2)$$

Earlier, however, we showed that, if the state vector transforms as

$$|\psi'\rangle = \hat{U}\,|\psi\rangle$$

then the operators transform as

$$\hat{U}^{-1}\hat{O}\hat{U}$$

which then implies that the rotation operator is

$$\hat{U}(\vec{\alpha}) = e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}$$

as we expect.

This result derived for infinitesimal rotation angles holds for finite rotation angles also.

Let us look at the effect on state vectors

$$\vec{r}\,'_{op}\,|\vec{r}_0\rangle = \vec{r}\,'_0\,|\vec{r}_0\rangle = e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\vec{r}_{op}e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle$$

$$\vec{r}_{op}e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle = \vec{r}\,'_0e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle$$

or

$$e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle$$

is an eigenstate of $\vec{r}_{op}$ with eigenvalue $\vec{r}\,'_0$ or

$$|\vec{r}\,'_0\rangle = e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle \text{ and } \langle\vec{r}\,'_0| = \langle\vec{r}_0|\,e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}$$

and thus $\vec{L}_{op}$ is the generator of ordinary rotations in 3-dimensional space.

For wave functions we have $\psi(\vec{r}_0) = \langle\vec{r}_0 \mid \psi\rangle$.

Using this result, the wave function transforms as

$$\psi(\vec{r}\,'_0) = \langle \vec{r}\,'_0 \mid \psi \rangle = \langle \vec{r}_0 | \, e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} \, | \psi \rangle$$
$$= \langle \vec{r}_0 \mid \psi' \rangle = \psi'(\vec{r}_0) = \text{ wave function at rotated point } \vec{r}\,'_0$$

Now, we have seen that

$$\vec{L}_{op} = \vec{r}_{op} \times \vec{p}_{op} = -i\hbar \vec{r}_{op} \times \nabla$$

This can easily be evaluated in different coordinate systems.

Cartesian Coordinates

$$\vec{L}_{op} = -i\hbar \sum_{ijk} \varepsilon_{ijk} x_j \frac{\partial}{\partial x_k} \hat{e}_i$$

$$\hat{L}_1 = -i\hbar \left(x_2 \frac{\partial}{\partial x_3} - x_3 \frac{\partial}{\partial x_2} \right)$$

$$\hat{L}_2 = -i\hbar \left(x_3 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_3} \right) \quad , \quad \hat{L}_3 = -i\hbar \left(x_1 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_1} \right)$$

as we saw at the beginning of this discussion.

Spherical-Polar Coordinates

We have $$\vec{r} = r\hat{e}_r \text{ and } \nabla = \hat{e}_r \frac{\partial}{\partial r} + \hat{e}_\theta \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{e}_\varphi \frac{1}{r \sin\theta} \frac{\partial}{\partial \varphi}$$

where

$$\hat{e}_r = \sin\theta \cos\varphi \hat{e}_1 + \sin\theta \sin\varphi \hat{e}_2 + \cos\theta \hat{e}_3$$
$$\hat{e}_\theta = \cos\theta \cos\varphi \hat{e}_1 + \cos\theta \sin\varphi \hat{e}_2 + \sin s\theta \hat{e}_3$$
$$\hat{e}_\varphi = -\sin\theta \hat{e}_1 + \cos\theta \hat{e}_2$$

which gives

$$\vec{L}_{op} = -i\hbar\left[\hat{e}_\varphi \frac{\partial}{\partial\theta} - \hat{e}_\theta \frac{1}{\sin\theta}\frac{\partial}{\partial\varphi}\right]$$

and

$$\hat{L}_3 = \hat{L}_z = \hat{e}_z \cdot \vec{L}_{op} = -i\hbar\frac{\partial}{\partial\varphi}$$

$$\vec{L}_{op}^2 = \vec{L}_{op}\cdot\vec{L}_{op} = -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right]$$

Similarly, we have

$$\hat{L}_1 = \hat{L}_x = i\hbar\left[\sin\varphi\frac{\partial}{\partial\theta} + \cot\theta\cos\varphi\frac{\partial}{\partial\varphi}\right]$$

$$\hat{L}_2 = \hat{L}_y = -i\hbar\left[\cos\varphi\frac{\partial}{\partial\theta} - \cot\theta\sin\varphi\frac{\partial}{\partial\varphi}\right]$$

and

$$\nabla^2 = \frac{1}{r}\frac{\partial^2}{\partial r^2}r - \frac{\vec{L}_{op}^2}{\hbar^2 r^2}$$

Eigenfunctions; Eigenvalues; Position Representation

Since $\left[\hat{L}_i, \hat{L}_j\right] = i\hbar\varepsilon_{ijk}\hat{L}_k$ and $\left[\hat{L}_{op}^2, \hat{L}_j\right] = 0$

the derivation of the eigenvalues and eigenvectors follows our earlier work, i.e., the equations

$$\hat{L}_{op}^2|\ell m\rangle = \hbar^2\ell(\ell+1)|\ell m\rangle \text{ and } \hat{L}_3|\ell m\rangle = \hbar m|\ell m\rangle$$

$$\hat{L}_\pm = \hat{L}_x \pm i\hat{L}_y$$

imply

$$\ell = \frac{\text{integer}}{2} \geq 0$$

and for a given value of ℓ , m takes on the $2\ell+1$ values

$$m = -\ell, -\ell+1, -\ell+2, \ldots\ldots\ldots, \ell-2, \ell-1, \ell$$

If we move back into 3-dimensional space and define

$$Y_{\ell m}(\theta,\varphi) = \langle \theta\varphi \mid \ell m \rangle = \text{ spherical harmonic}$$

then we have the defining equations for the $Y_{\ell m}(\theta,\phi)$ given by

$$\begin{aligned}
\langle \theta\varphi | \vec{L}^2_{op} | \ell m \rangle &= \vec{L}^2_{op} \langle \theta\varphi \mid \ell m \rangle = \vec{L}^2_{op} Y_{\ell m}(\theta,\varphi) \\
&= \hbar^2 \ell(\ell+1) \langle \theta\varphi \mid \ell m \rangle = \hbar^2 \ell(\ell+1) Y_{\ell m}(\theta,\varphi) \\
\langle \theta\varphi | \hat{L}_3 | \ell m \rangle &= \hat{L}_3 \langle \theta\varphi \mid \ell m \rangle = \hat{L}_3 Y_{\ell m}(\theta,\varphi) \\
&= \hbar m \langle \theta\varphi \mid \ell m \rangle = \hbar m Y_{\ell m}(\theta,\varphi)
\end{aligned}$$

Before determining the functional form of the $Y_{\ell m}{}'$s, we must step back and see if there are any restrictions that need to be imposed on the possible eigenvalues ℓ and m due to the fact that we are in a real 3-dimensional space.

In general, eigenvalue restrictions come about only from the imposition of physical boundary conditions.

A standard boundary condition that is usually imposed is the following:

In real 3-dimensional space, if we rotate the system (or the axes by 2π, then we get back the same world.

This means that the $Y_{\ell m}(\theta,\phi)$ should be single-valued under such rotations or that

$$Y_{\ell m}(\theta,\varphi+2\pi) = Y_{\ell m}(\theta,\varphi)$$

Now from the general rules we have developed, this gives

$$\langle \theta, \varphi + 2\pi \mid \ell m \rangle = \langle \theta\varphi | \, e^{2\pi i \hat{L}_3/\hbar} \, | \ell m \rangle = e^{2\pi m i} \langle \theta, \varphi \mid \ell m \rangle$$

or that single-valuedness of the wave function requires that

$$\langle \theta, \varphi + 2\pi \mid \ell m \rangle = \langle \theta\varphi | \, e^{2\pi i \hat{L}_3/\hbar} \, | \ell m \rangle = e^{2\pi m i} \langle \theta, \varphi \mid \ell m \rangle$$

This says, that for orbital angular momentum in real 3-dimensional space, no 1/2-integer values are allowed for *m* and hence that ℓ must be an integer.

The allowed sets are:

$$\ell = 0 \quad , \quad m = 0$$
$$\ell = 1 \quad , \quad m = -1, 0, 1$$
$$\ell = 2 \quad , \quad m = -2, -1, 0, 1, 2$$

and so on.

It is important to note that we are imposing a much stronger condition than is necessary.

In general, as we have stated several times, it is not the state vectors, operators or wave functions that have any physical meaning in quantum mechanics.

The only quantities that have physical meaning are those directly related to measurable quantities, namely, the probabilities, and the expectation values.

This single-valued condition on m is not needed for the single-valuedness of the expectation values, since the extra phase factors involving *m* will cancel out during the calculation of the expectation value.

Experiment, however, says that ℓ is an integer only, so it seems that the strong condition is valid.

We cannot prove that this is so, however.

Let us now figure out the $Y_{\ell m}(\theta,\phi)$.

We have

$$\hat{L}_3 Y_{\ell m}(\theta,\varphi) = \frac{\hbar}{i}\frac{\partial}{\partial\varphi} Y_{\ell m}(\theta,\varphi) = \hbar m Y_{\ell m}(\theta,\varphi)$$

which tells us the ϕ-dependence.

$$Y_{\ell m}(\theta,\varphi) = e^{im\varphi} P_{\ell m}(\theta)$$

Now we also must have, since ℓ = maximum value of m

$$\langle\theta\varphi|\,\hat{L}_+\,|\ell\ell\rangle = 0 = \hat{L}_+\,\langle\theta\varphi\,|\,\ell\ell\rangle = \hat{L}_+ Y_{\ell\ell}(\theta,\varphi)$$

Using the expressions for $\hat{L}_x, \hat{L}_y, \hat{L}_+$ and $\hat{L}_-$ we have

$$\frac{\hbar}{i} e^{i\varphi}\left[i\frac{\partial}{\partial\theta} - \cot\theta\frac{\partial}{\partial\varphi}\right] Y_{\ell\ell}(\theta,\varphi) = 0$$

Using

$$\frac{\hbar}{i}\frac{\partial}{\partial\varphi} Y_{\ell\ell}(\theta,\varphi) = \ell\hbar Y_{\ell\ell}(\theta,\varphi)$$

we get

$$\left[\frac{\partial}{\partial\theta} - \ell\cot\theta\right] P_{\ell\ell}(\theta) = 0$$

or

$$P_{\ell\ell}(\theta) = (\sin\theta)^\ell$$

Therefore, the final expression is

$$Y_{\ell\ell}(\theta,\varphi) = A_{\ell m} e^{i\ell\varphi}(\sin\theta)^\ell$$

Now that we have generated the topmost spherical harmonic.

We can generate all the others for a given ℓ using the lowering operators, i.e.,

$$\hat{L}_- Y_{\ell m}(\theta,\varphi) = \hbar\sqrt{\ell(\ell+1)-m(m-1)}\,Y_{\ell m-1}(\theta,\varphi)$$

where

$$\hat{L}_- = \frac{\hbar}{i}e^{-i\varphi}\left[i\frac{\partial}{\partial\theta} - \cot\theta\frac{\partial}{\partial\varphi}\right]$$

In general, we choose the $A_{\ell m}$ so that the $Y_{\ell m}$ (θ,ϕ) are normalized, i.e.,

$$1 = \int d\Omega\,|Y_{\ell m}(\theta,\varphi)|^2 = \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta d\theta\,|Y_{\ell m}(\theta,\varphi)|^2$$

Since the $Y_{\ell m}$ (θ,ϕ) are eigenfunctions of a Hermitian operators, they form a compete set which we can always make orthogonal so that we also assume that we have

$$\int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta d\theta Y^*_{\ell' m'}(\theta,\varphi)Y_{\ell m}(\theta,\varphi) = \delta_{\ell'\ell}\delta_{m'm}$$

The algebra is complicated.

The general result is

$$Y_{\ell m}(\theta,\varphi) - \frac{(-1)\ell}{2^\ell \ell!}\sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell+m)!}{(\ell-m)!}}\frac{e^{im\varphi}}{(\sin\theta)^m}\left(\frac{d}{d\cos\theta}\right)^{\ell-m}(\sin\theta)^{2\ell}$$

Some examples are:

$$Y_{00} = \frac{1}{\sqrt{4\pi}} \quad , \quad Y_{10} = \sqrt{\frac{3}{4\pi}} \cos\theta$$

$$Y_{1,\pm 1} = \mp\sqrt{\frac{3}{8\pi}} e^{\pm i\varphi} \sin\theta \quad , \quad Y_{20} = \sqrt{\frac{5}{16\pi}} (3\cos^2\theta - 1)$$

$$Y_{2,\pm 1} = \mp\sqrt{\frac{15}{8\pi}} \sin\theta \cos\theta e^{\pm i\varphi} \quad , \quad Y_{2,\pm 2} = \sqrt{\frac{15}{32\pi}} \sin^2\theta e^{\pm 2i\varphi}$$

Some Properties

$$Y_{\ell,-m}(\theta,\varphi) = (-1)^m Y^*_{\ell,m}(\theta,\varphi)$$

Under the parity operation

$$\vec{r} \to -\vec{r} \text{ or } r \to r, \theta \to \pi - \theta, \varphi \to \varphi + \pi$$

This says that

$$e^{im\varphi} \to e^{im\varphi} e^{im\pi} = (-1)^m e^{im\varphi}$$

$$\sin\theta \to \sin(\pi - \theta) \to \sin\theta$$

$$\cos\theta \to \cos(\pi - \theta) \to -\cos\theta$$

which imply that

$$Y_{\ell,m}(\theta,\varphi) \to (-1)^\ell Y_{\ell,m}(\theta,\varphi)$$

Therefore

if ℓ is even, then we have an even parity state

if ℓ is odd, then we have an odd parity state

Since they form a complete set, any function of (θ, ϕ) can be expanded in terms of the $Y_{\ell m}(\theta,\phi)$ (the $Y_{\ell m}(\theta,\phi)$ are a basis), i.e., we can write

$$f(\theta,\varphi) = \sum_{\ell,m} f_{\ell m} Y_{\ell,m}(\theta,\varphi)$$

where

$$f_{\ell m} = \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta d\theta Y^*_{\ell' m'}(\theta,\varphi) f(\theta,\varphi)$$

Spin

As we discussed earlier, a second kind of angular momentum exists in quantum mechanics.

It is related to internal degrees of freedom of particles and is not related in any way to ordinary 3-dimensional space properties.

We designate this new angular momentum by spin and represent it by the operator $\vec{S}_{op}$, where since it is an angular momentum, its components must satisfy the commutators

$$\left[\hat{S}_i, \hat{S}_j\right] = i\hbar\varepsilon_{ijk}\hat{S}_k$$

where we have used the Einstein summation convention over repeated indices.

The analysis for the eigenvalues and eigenvectors follows the same path as for earlier discussions.

We have

$$\vec{S}^2_{op}\left|s, m_s\right\rangle = \hbar^2 s(s+1)\left|s, m_s\right\rangle$$
$$\hat{S}_3\left|s, m_s\right\rangle = \hbar m_s\left|s, m_s\right\rangle$$

which, together with the commutators, gives the following results.

For a given value of s, we have $2s + 1$ m_s-values

$$m_s = -s, -s+1, -s+2, \ldots\ldots, s-2, s-1, s$$

where

$$s = \frac{\text{integer}}{2} \geq 0$$

There are no boundary conditions restricting the value of s, so we can have both integer and half-integer values.

We now turn our attention to a most important special case and then generalize the details.

Spin 1/2

We define a new operator $\vec{\sigma}_{op}$ such that

$$\vec{S}_{op} = \frac{1}{2}\hbar\vec{\sigma}_{op}$$

where $\vec{\sigma}_{op} = (\ \hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$ are called the Pauli spin operators.

It is experimentally observed that if one measures the component of this spin angular momentum along any direction, one always obtains either $\pm\hbar/2$.

If we designate the state with spin = $+\hbar$ /2 or spin up in the $\hat{n}$ direction by the ket vectors $|\ \hat{n}\ \uparrow\rangle$ or $|\ \hat{n}\ +\rangle$, we then have

$$\vec{S}_{op} \cdot \hat{n}\,|\hat{n}\uparrow\rangle = \frac{\hbar}{2}\,|\hat{n}\uparrow\rangle \text{ and } \vec{S}_{op} \cdot \hat{n}\,|\hat{n}\downarrow\rangle = -\frac{\hbar}{2}\,|\hat{n}\downarrow\rangle$$

Any pair of eigenvectors $|\ \hat{n}\uparrow\rangle$ or $|\ \hat{n}\downarrow\rangle$ for a given direction $\hat{n}$ form a basis for the vector space associated with spin = 1/2 and we have

$\langle \hat{n} \uparrow | \psi \rangle =$ amplitude for finding spin "up" along $\hat{n}$ if we are in the state $|\psi\rangle$

$\langle \hat{n} \downarrow | \psi \rangle =$ amplitude for finding spin "down" along $\hat{n}$ if we are in the state $|\psi\rangle$

These two amplitudes exhaust all possible measurements for spin = 1/2 in the $\hat{n}$ direction and therefore completely specify the state $|\psi\rangle$.

That is what we mean physically when we say they form a basis set or representation.

When we build the standard basis from these amplitudes, we choose it to be an eigenstate of the $\hat{S}_3$ or $\hat{S}_z$ operator, i.e., if we write

$$|\psi\rangle = \begin{pmatrix} \langle z \uparrow | \psi \rangle \\ \langle z \downarrow | \psi \rangle \end{pmatrix} = \text{a 2 - component vector}$$

then the appropriate basis is

$$|z \uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \text{spin up in z - direction}$$

$$|z \downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \text{spin down in z - direction}$$

Matrix Representations

Using this basis, the matrix representation of $\hat{S}_z = \frac{\hbar}{2}\hat{\sigma}_z$ is

$$S_z = \begin{pmatrix} \langle z \uparrow | \hat{S}_z | z \uparrow \rangle & \langle z \uparrow | \hat{S}_z | z \downarrow \rangle \\ \langle z \downarrow | \hat{S}_z | z \uparrow \rangle & \langle z \downarrow | \hat{S}_z | z \downarrow \rangle \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\hbar}{2}\sigma_z$$

$$\rightarrow \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Now

$$\hat{S}_{\pm} = \hat{S}_x \pm i\hat{S}_y$$

$$\rightarrow \hat{S}_x = \frac{\hat{S}_+ + \hat{S}_-}{2} \text{ and } \hat{S}_y = \frac{\hat{S}_+ - \hat{S}_-}{2i}$$

Therefore, using

$$S_x = \begin{pmatrix} \langle z\uparrow|\hat{S}_x|z\uparrow\rangle & \langle z\uparrow|\hat{S}_x|z\downarrow\rangle \\ \langle z\downarrow|\hat{S}_x|z\uparrow\rangle & \langle z\downarrow|\hat{S}_x|z\downarrow\rangle \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\hbar}{2}\sigma_x$$

$$\rightarrow \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and in a similar way

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

Properties of the $\hat{\sigma}_i$

From

$$\left[\hat{S}_i, \hat{S}_j\right] = i\hbar\varepsilon_{ijk}\hat{S}_k$$

and

$$\vec{S}_{op} = \frac{1}{2}\hbar\vec{\sigma}_{op}$$

we get the commutation relations

$$\left[\hat{\sigma}_i, \hat{\sigma}_j\right] = 2i\varepsilon_{ijk}\hat{\sigma}_k$$

In addition, we have

$$\hat{\sigma}_i\hat{\sigma}_j = i\varepsilon_{ijk}\hat{\sigma}_k \quad i \neq j$$

$$\hat{\sigma}_i\hat{\sigma}_j + \hat{\sigma}_j\hat{\sigma}_i = \{\hat{\sigma}_i, \hat{\sigma}_j\} = 0 \text{ (called the anticommutator)}$$

$$\hat{\sigma}_{\mathrm{i}}^2 = \hat{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The fact that the spin = 1/2 operators anticommute is directly linked to the existence of fermions as sees when the relativistic equation for the electron is studied.

Put all together, these relations give

$$\hat{\sigma}_i\hat{\sigma}_j = \delta_{ij} + i\varepsilon_{ijk}\hat{\sigma}_k$$

or going back to the $\hat{S}_i$

$$\hat{S}_i\hat{S}_j = \frac{\hbar^2}{4}\delta_{ij} + i\varepsilon_{ijk}\frac{\hbar}{2}\hat{S}_k$$

In the special case of spin = 1/2, we have

$$\vec{S}_{op}^2 = \frac{\hbar^2}{4}\left(\hat{\sigma}_x^2 + \hat{\sigma}_y^2 + \hat{\sigma}_z^2\right) = \frac{\hbar^2}{4}\left(\hat{I} + \hat{I} + \hat{I}\right)$$

$$= \frac{3\hbar^2}{4}\hat{I} = \hbar^2 s(s+1)\hat{I} \text{ with } s = \frac{1}{2}$$

A very useful property that we will employ many time later on is

$$\begin{aligned}(\vec{a}\cdot\vec{\sigma}_{op})\left(\vec{b}\cdot\vec{\sigma}_{op}\right) &= a_i\hat{\sigma}_i b_j\hat{\sigma}_j \\ &= \delta_{ij}a_ib_j + i\varepsilon_{ijk}a_ib_j\hat{\sigma}_k = a_ib_i + i\varepsilon_{ijk}a_ib_j\hat{\sigma}_k \\ &= \vec{a}\cdot\vec{b} + i(\vec{a}\times\vec{b})\cdot\vec{\sigma}_{op}\end{aligned}$$

Rotations in Spin Space

We have said that

$$\vec{S}_{op}\cdot\hat{n}\,|\hat{n}\pm\rangle = \pm\frac{\hbar}{2}\,|\hat{n}\pm\rangle$$

What do the states $|\,\hat{n}\,\pm\rangle$ look like in the $|\,\hat{z}\,\pm\rangle$ basis?

One way to find out is the direct approach.

Let us choose a Cartesian basis for the unit vector (we will look at other choices afterwards)

$$\hat{n} = (n_x, n_y, n_z) \quad , \qquad \text{all real}$$

We then have

$$\vec{S}_{op}\cdot\hat{n} = \frac{\hbar}{2}\vec{\sigma}_{op}\cdot\hat{n} = \frac{\hbar}{2}(n_x\hat{\sigma}_x + n_y\hat{\sigma}_y + n_z\hat{\sigma}_z)$$

$$= \frac{\hbar}{2}\left(n_x\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} + n_y\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix} + n_z\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\right)$$

$$= \frac{\hbar}{2}\begin{pmatrix} n_z & n_x - in_y\\ n_x + in_y & -n_z\end{pmatrix}$$

and

$$\vec{S}_{op}\cdot\hat{n}\,|\hat{n}+\rangle = +\frac{\hbar}{2}\,|\hat{n}+\rangle$$

$$\frac{\hbar}{2}\begin{pmatrix} n_z & n_x - in_y\\ n_x + in_y & -n_z\end{pmatrix}\begin{pmatrix} a\\ b\end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} a\\ b\end{pmatrix}$$

where we have represented

$$|\hat{n}+\rangle = \begin{pmatrix} a\\ b\end{pmatrix}$$

This matrix equation gives two homogeneous equations for a and b

$$(n_z - 1)a + (n_x - in_y)b = 0$$
$$(n_x + in_y)a - (n_z + 1)b = 0$$

or

$$\frac{a}{b} = -\frac{n_x - in_y}{n_z - 1} = \frac{n_z + 1}{n_x + in_y}$$

The homogeneous equations have a non-trivial solution only if the determinant of the coefficients of a and b equals zero or $(n_z + 1)(n_z - 1) + (n_x - in_y)(n_x + in_y) = 0$

We assume that the vector $|\ \hat{n}\ +\rangle$ is normalized to one or

$$|a|^2 + |b|^2 = 1$$

Putting all this together we get $a = \frac{1}{\sqrt{2}}\sqrt{1 + n_z}$ and $b = \frac{1}{\sqrt{2}}\sqrt{1 - n_z}$

$$|\hat{n}+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{1 + n_z} \\ \sqrt{1 - n_z} \end{pmatrix}$$

We can easily check this by letting $|\quad +\rangle = |\quad +\rangle$ or $n_z = 1, n_x = n_y = 0$ which gives

$$|\hat{z}+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

as expected.

In a similar manner

$$|\hat{n}-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} -\sqrt{1 - n_z} \\ \sqrt{1 + n_z} \end{pmatrix}$$

Note that the two vectors $|\ \hat{n}\ +\rangle$ and $|\ \hat{n}\ -\rangle$ are orthonormal as we expect.

This calculation can also be carried out in other coordinate bases.

For the spherical-polar basis we get

$$\hat{n} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$$

$$\vec{S}_{op}\cdot\hat{n} = \frac{\hbar}{2}\begin{pmatrix} \cos\theta & e^{-i\varphi}\sin\theta \\ e^{i\varphi}\sin\theta & -\cos\theta \end{pmatrix}$$

$$|\hat{n}+\rangle = \begin{pmatrix} \cos\frac{\theta}{2} \\ e^{i\varphi}\sin\frac{\theta}{2} \end{pmatrix} \text{ and } |\hat{n}-\rangle = \begin{pmatrix} \sin\frac{\theta}{2} \\ -e^{i\varphi}\cos\frac{\theta}{2} \end{pmatrix}$$

What can we say about operators in the spin = 1/2 vector space?

Any such operator $\hat{B}$ can be expressed as a linear combination of the four linearly independent matrices $\{\hat{I}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}$ (they are, in fact, a basis for all 2 × 2 matrices)

$$\hat{B} = a_0\hat{I} + a_x\hat{\sigma}_x + a_y\hat{\sigma}_y + a_z\hat{\sigma}_z = a_0\hat{I} + a\cdot\vec{\sigma}_{op} = \begin{pmatrix} a_0 + a_z & a_x - ia_y \\ a_x + ia_y & a_0 - a_z \end{pmatrix}$$

In particular, the density operator or the state operator $\hat{W}$ which is the same as any other operator can be written as

$$\hat{W} = \frac{1}{2}(\hat{I} + \vec{a}\cdot\vec{\sigma}_{op})$$

where the factor of 1/2 has been chosen so that we have the required property

$$Tr\hat{W} = \frac{1}{2}Tr\hat{I} + a_i Tr\sigma_i = 1 \text{ since } Tr\sigma_i = 0$$

Since $\hat{W} = \hat{W}^{\dagger}$ we must also have all the a_i real.

What is the physical meaning of the vector $\vec{a}$?

Consider the following

$$\begin{aligned}\langle\hat{\sigma}_x\rangle &= Tr(\hat{W}\hat{\sigma}_x) = \frac{1}{2}Tr((\hat{I} + a_x\hat{\sigma}_x + a_y\hat{\sigma}_y + a_z\hat{\sigma}_z)\hat{\sigma}_x) \\ &= \frac{1}{2}Tr(\hat{\sigma}_x + a_x\hat{\sigma}_x^2 + a_y\hat{\sigma}_y\hat{\sigma}_x + a_z\hat{\sigma}_z\hat{\sigma}_x) \\ &= \frac{1}{2}Tr(\hat{\sigma}_x + a_x\hat{I} - ia_y\hat{\sigma}_z + ia_z\hat{\sigma}_y) \\ &= \frac{a_x}{2}Tr(\hat{I}) = a_x\end{aligned}$$

or, in general,

$$\langle\vec{\sigma}_{op}\rangle = Tr\left(\hat{W}\vec{\sigma}_{op}\right) = \vec{a} = \text{ polarization vector}$$

Now the eigenvalues of $\hat{W}$ are equal to

$$\frac{1}{2} + \text{ eigenvalues of } \vec{a}\cdot\vec{\sigma}_{op}$$

and from our earlier work the eigenvalues are $\vec{a}\cdot\vec{\sigma}_{op} = \pm 1$.

Therefore the eigenvalues of $\hat{W}$ are

$$\frac{1}{2}(1 \pm |\vec{a}|)$$

But, as we showed earlier, all eigenvalues of $\hat{W}$ are ≥ 0, which says that then polarization vectors have a length $|\vec{a}|$ restricted to $0 \leq |\vec{a}| \leq 1$.

Pure states have $|\vec{a}| = 1$ and this gives eigenvalues 1 and 0 for $\hat{W}$, which corresponds to *maximum* polarization.

Note that for $\vec{a} = \hat{e}_3$ a $= \hat{e}_3$, we have

$$\hat{W} = \frac{1}{2}(\hat{I}+\hat{\sigma}_3) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = |z+\rangle\langle z+|$$

as it should for a pure state.

An unpolarized state has $|\vec{a}| = 0$ and this gives eigenvalues (1/2, 1/2) for $\hat{W}$.

This represents an *isotropic* state where $\langle \hat{S}_i \rangle = 0$.

In this case we have

$$\hat{W} = \frac{1}{2}\hat{I} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{2}|z+\rangle\langle z+| + \frac{1}{2}|z-\rangle\langle z-|$$

as we expect for a nonpure state or a mixture.

Let us now connect all of this stuff to rotations in spin space.

For an infinitesimal rotation through angle $\alpha = |\vec{\alpha}|$ about an axis along $\alpha\hat{} = \vec{\alpha}/\alpha$, a unit vector $\hat{m}$ becomes

$$\hat{n} = \hat{m} + \vec{\alpha} \times \hat{m}$$

Following the same steps as earlier, this implies

$$\vec{S}_{op} \cdot \hat{n} = \vec{S}_{op} \cdot \hat{m} + \vec{S}_{op} \cdot (\vec{\alpha} \times \hat{m}) = \vec{S}_{op} \cdot \hat{m} + \varepsilon_{ijk}\alpha_i \hat{m}_j \hat{S}_k$$

But we have

$$\varepsilon_{ijk}\hat{S}_k = \frac{1}{i\hbar}\left[\hat{S}_i, \hat{S}_j\right]$$

which implies that

$$\vec{S}_{op}\cdot\hat{n} = \vec{S}_{op}\cdot\hat{m} + \frac{1}{i\hbar}\left[\hat{S}_i, \hat{S}_j\right]\alpha_i\hat{m}_j$$

$$= \vec{S}_{op}\cdot\hat{m} + \frac{1}{i\hbar}[\vec{S}_{op}\cdot\vec{\alpha}, \vec{S}_{op}\cdot\hat{m}]$$

Using the same approximations as we did earlier, we can see that this expression, to first order in α, is equivalent to

$$\vec{S}_{op}\cdot\hat{n} = e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}}\vec{S}_{op}\cdot\hat{m}e^{\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}}$$

This result holds for finite rotation angles also.

Now using this result, we have

$$\vec{S}_{op}\cdot\hat{n}\left[e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}}|\hat{m}+\rangle\right] = \left[e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}}\vec{S}_{op}\cdot\hat{m}\right]|\hat{m}+\rangle$$

$$= \frac{\hbar}{2}\left[e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}}|\hat{m}+\rangle\right]$$

This says that

$$e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}}|\hat{m}+\rangle = |\hat{n}+\rangle \text{ and similarly } e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}}|\hat{m}-\rangle = |\hat{n}-\rangle$$

The rotation that takes $\hat{m} \rightarrow \hat{n}$ is not unique, however.

We are free to rotate by an arbitrary amount about $\hat{n}$ after rotating $\hat{m}$ into $\hat{n}$.

This freedom corresponds to adding a phase factor.

We say that the unitary operator

$$e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}}$$

has the effect of rotating the eigenstate of $\vec{S}_{op}\cdot\hat{m}$ into the eigenstate of $\vec{S}_{op}\cdot\hat{n}$.

The operator performs rotations on the spin degrees of freedom.

The equations are analogous to those for real rotations in space generated by $\vec{L}_{op}$.

Let us now work out a very useful identity.

We can write

$$e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}} = e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}} = \sum_{n=0}^{\infty}\frac{\left(-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}\right)^n}{n!}$$

Now we have

$$\left(\vec{\sigma}_{op}\cdot\vec{\alpha}\right)^2 = \vec{\alpha}\cdot\vec{\alpha} + i(\vec{\alpha}\times\vec{\alpha})\cdot\vec{\sigma}_{op} = \vec{\alpha}^2 = \alpha^2$$

Therefore

$$\begin{aligned} e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}} &= \hat{I} + \frac{1}{1!}\left(-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}\right) + \frac{1}{2!}\left(-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}\right)^2 + \frac{1}{3!}\left(-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}\right)^3 + \ldots \\ &= \hat{I}\left(1 - \frac{\left(\frac{\alpha}{2}\right)^2}{2!} + \frac{\left(\frac{\alpha}{2}\right)^4}{4!} - \ldots.\right) - i\vec{\sigma}_{op}\cdot\hat{\alpha}\left(\frac{\left(\frac{\alpha}{2}\right)}{1!} - \frac{\left(\frac{\alpha}{2}\right)^3}{3!} + \ldots.\right) \\ &= \hat{I}\cos\frac{\alpha}{2} - i(\vec{\sigma}_{op}\cdot\hat{\alpha})\sin\frac{\alpha}{2} \end{aligned}$$

Consider an example.

$\vec{\alpha} \rightarrow -90^\circ$ rotation about the *x*-axis or

$$\vec{\alpha} \rightarrow -\frac{\pi}{2}\hat{x}$$

Now we have

$$\hat{\sigma}_z(\hat{\sigma}_x)^n = (-\hat{\sigma}_x)^n\hat{\sigma}_z$$

Using these relations we get

$$e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}}\hat{\sigma}_z e^{\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}} = e^{\frac{i\pi}{4}\hat{\sigma}_x}\hat{\sigma}_z e^{-\frac{i\pi}{4}\hat{\sigma}_x} = e^{\frac{i\pi}{4}\hat{\sigma}_x} e^{\frac{i\pi}{4}\hat{\sigma}_x}\hat{\sigma}_z = e^{\frac{i\pi}{2}\hat{\sigma}_x}\hat{\sigma}_z$$

$$= (\cos\frac{\pi}{2} + i\hat{\sigma}_x \sin\frac{\pi}{2})\hat{\sigma}_z = i\hat{\sigma}_x\hat{\sigma}_z = \hat{\sigma}_y$$

as expected for this particular rotation.

Now the spin degrees of freedom are simply additional degrees of freedom for the system.

The spin degrees of freedom are independent of the spatial degrees of freedom, however.

This means that we can specify them independently or that $\vec{S}_{op}$ commutes with all operators that depend on 3-dimensional space

$$\left[\vec{S}_{op}, \vec{r}_{op}\right] = 0 \quad , \quad \left[\vec{S}_{op}, \vec{p}_{op}\right] = 0 \quad , \quad \left[\vec{S}_{op}, \vec{L}_{op}\right] = 0$$

To specify completely the state of a *spinning* particle or a system with internal degrees of freedom of any kind we must know

1. the amplitudes for finding the particle at points in space
2. the amplitudes for different spin orientations

By convention, we choose $\hat{z}$ as the spin quantization direction for describing state vectors.

The total state vector is then a direct-product state of the form

$$|\psi\rangle = |\text{external}\rangle \otimes |\text{internal}\rangle$$

and the amplitudes are

$\langle \vec{r}, \hat{z}+ \mid \psi \rangle =$ probability amplitude for finding the particle at $\vec{r}$ with spin up in the $\hat{z}$ direction

$\langle \vec{r}, \hat{z}- \mid \psi \rangle =$ probability amplitude for finding the particle at $\vec{r}$ with spin down in the $\hat{z}$ direction

where

$$\langle \vec{r}, \hat{z}+ \mid \psi \rangle = \langle \vec{r} \mid \psi_{external} \rangle \langle \hat{z}+ \mid \psi_{\text{internal}} \rangle$$

and so on.

The total probability density for finding a particle at $\vec{r}$ is then the sum of two terms

$$|\langle \vec{r}, \hat{z}+ \mid \psi \rangle|^2 + |\langle \vec{r}, \hat{z}- \mid \psi \rangle|^2$$

which represents a sum over all the ways of doing it.

The total angular momentum of such a *spinning* particle is the sum of its orbital and spin angular momenta

$$\vec{J}_{op} = \vec{L}_{op} + \vec{S}_{op}$$

$\vec{J}_{op}$ is now the generator of rotations in 3-dimensional space and in spin space or it affects both external and internal degrees of freedom.

If we operate with $e^{-i\vec{\alpha}\cdot\vec{J}_{op}/\hbar}$ on the basis state $|\vec{r}_0, \hat{m}+\rangle$, where the particle is definitely at $\vec{r}_0$ with spin up (+ $\hbar$ /2) in the $\hat{m}$ direction, then we get a new state $|\vec{r}'_0, \hat{n}+\rangle$ where

$$\vec{r}\,'_0 = \vec{r}_0 + \vec{\alpha} \times \vec{r}_0 \text{ and } \hat{n} = \hat{m} + \vec{\alpha} \times \hat{m}$$

Since $[\vec{S}_{op}, \vec{L}_{op}] = 0$ we have

$$e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{J}_{op}} = e^{-\frac{i}{\hbar}(\vec{\alpha}\cdot\vec{L}_{op}+\vec{\alpha}\cdot\vec{S}_{op})}$$

$$= e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{S}_{op}} e^{-\frac{i}{\hbar}\vec{\alpha}\cdot[\vec{S}_{op},\vec{L}_{op}]}$$

$$= e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{S}_{op}}$$

This implies that

$$e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{J}_{op}} \left|\vec{r}_0, \hat{m}+\right\rangle = e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{S}_{op}} \left|\vec{r}_0, \hat{m}+\right\rangle = \left|\vec{r}\,'_0, \hat{n}+\right\rangle$$

and that $e^{-i\vec{\alpha}\cdot\vec{L}_{op}/\hbar}$ carries out the rotation of the spatial degrees of freedom, while $e^{-i\vec{\alpha}\cdot\vec{S}_{op}/\hbar}$ carries out the rotation of the spin degrees of freedom.

If

$$\left|\psi'\right\rangle = e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{J}_{op}} \left|\psi\right\rangle$$

then the wave function of $|\psi'\rangle$ is

$$\left\langle\vec{r}_0, \hat{m}+ \mid \psi'\right\rangle = \left\langle\vec{r}_0, \hat{m}+\right| e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{J}_{op}} \left|\psi\right\rangle = \left\langle\vec{r}\,'_0, \hat{n}+ \mid \psi\right\rangle$$

This is the wavefunction of $|\psi\rangle$ evaluated at the rotated point with a rotated spin quantization direction.

The *spin representation* of rotations has a feature which is strikingly different from that of rotations in 3-dimensional space.

Consider a rotation by 2π in spin space.

This implies that

$$e^{-i\pi\vec{\sigma}_{op}\cdot\vec{\alpha}} = \cos\pi - i\vec{\sigma}_{op}\cdot\vec{\alpha}\sin\pi = -\hat{I}$$

A 2π rotation in spin space is represented by $-\hat{I}$.

Since the rotations $\vec{\alpha}$ and $\vec{\alpha} + 2\pi\vec{\alpha}$ are physically equivalent, we must say that the spin representation of rotations is *double-valued*, i.e.,

$$e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}} \text{ and } e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot(\vec{\alpha}+2\pi\hat{\alpha})} = -e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}}$$

represent the same rotation.

Superselection Rules

Let us expand on this important point.

The 2π rotation transformation operator is given by

$$\hat{U}_{\hat{n}}(2\pi) = e^{-\frac{2\pi i}{\hbar}\hat{n}\cdot\vec{J}_{op}}$$

When operating on the angular momentum state vectors we have

where $\hat{A}$ is any physical observable.

$$\hat{U}_{\hat{n}}(2\pi)\,|j,m\rangle = e^{-\frac{2\pi i}{\hbar}\hat{n}\cdot\vec{J}_{op}}\,|j,m\rangle = e^{-2\pi ij}\,|j,m\rangle = (-1)^{2j}\,|j,m\rangle$$

This says that it has no effect if j = integer and multiplies by -1 if j = half- integer.

We usually think of a rotation through 2π as a trivial operation that changes nothing in a physical system.

This belief implies that we are assuming all dynamical variables are invariant under 2π rotations or that

$$\hat{U}_{\hat{n}}(2\pi)\hat{A}\hat{U}^{\dagger}_{\hat{n}}(2\pi) = \hat{A}$$

where $\hat{A}$ is any physical observable.

But, as we have seen above, $\hat{U}_{\hat{n}}(2\pi)$ is not equal to a trivial operator (not equal to the identity operator for all physical states and operators).

This says that invariance under $\hat{U}_{\hat{n}}(2\pi)$ may lead to nontrivial consequences.

The consequences that arise from invariance of an observable are not identical to those that arise from invariance of a state.

Let $\hat{U}$ = a unitary operator that leaves the observable $\hat{F}$ invariant or that we have

$$\left[\hat{U}, \hat{F}\right] = 0$$

Now consider a state that is not invariant under the transformation $\hat{U}$.

If it is a pure state represented by $|\psi\rangle$, then $|\psi'\rangle = \hat{U}\,|\psi\rangle \neq |\psi\rangle$.

The expectation value of $\hat{F}$ in the state $|\psi\rangle$ is

$$\left\langle \hat{F} \right\rangle = \langle \psi' | \hat{F} | \psi' \rangle = \langle \psi | \hat{U}^{+} \hat{F} \hat{U} | \psi \rangle = \langle \psi | \hat{U}^{+} \hat{U} \hat{F} | \psi \rangle = \langle \psi | \hat{F} | \psi \rangle$$

which implies that the observable statistical properties of $\hat{F}$ are the same in the two states $|\psi\rangle$ and $|\psi'\rangle$.

This conclusion certainly holds for $\hat{U}(2\pi)$.

Does anything else hold?

Is there something peculiar to $\hat{U}(2\pi)$?

It turns out that $\hat{U}(2\pi)$ divides the vector space into two subspaces:

1. integer angular momentum - has states $|+\rangle$ where

$$\hat{U}(2\pi)\,|+\rangle = |+\rangle$$

2. half-integer angular momentum - has states $|-\rangle$ where

$$\hat{U}(2\pi)\,|-\rangle = -\,|-\rangle$$

Now for any invariant physical observable $\hat{B}$ (where $[\ \hat{U}, \hat{B}\ ] = 0$), we have

$$\langle +|\,\hat{U}(2\pi)\hat{B}\,|-\rangle = \langle +|\,\hat{B}\hat{U}(2\pi)\,|-\rangle$$
$$\langle +|\,\hat{B}\,|-\rangle = -\,\langle +|\,\hat{B}\,|-\rangle$$
$$\rightarrow \langle +|\,\hat{B}\,|-\rangle = 0$$

This says that all physical observable have vanishing matrix elements between states with integer angular momentum and states with half-integer angular momentum (states in the two subspaces).

This is called a **superselection rule**.

A superselection rule says that there is *no observable distinction* among vectors of the form

$$|\psi_\varphi\rangle = |+\rangle + e^{i\varphi}\,|-\rangle$$

for different values of the phase ϕ.

This is so because

$$\langle \psi_\varphi|\,\hat{B}\,|\psi_\varphi\rangle = \langle +|\,\hat{B}\,|+\rangle + \langle -|\,\hat{B}\,|-\rangle$$

and this is independent of ϕ.

What about a more general state represented by the state operator

$$\hat{W} = \sum_{i,j} w_{ij} \left|i\right\rangle \left\langle j\right|$$

If we break up the basis for the space into + and – subspaces, i.e.,

$$\text{basis set} = \{\{+ \text{ states}\}, \{- \text{ states}\}\}$$

then the matrix representation of $\hat{W}$ partitions into four blocks

$$\hat{W} = \begin{pmatrix} \hat{W}_{++} & \hat{W}_{+-} \\ \hat{W}_{-+} & \hat{W}_{--} \end{pmatrix}$$

While, for any physical observable, the same partitioning scheme produces

$$\hat{B} = \begin{pmatrix} \hat{B}_{++} & 0 \\ 0 & \hat{B}_{--} \end{pmatrix}$$

i.e., there is no mixing between the subspaces.

This gives

$$\left\langle \hat{B} \right\rangle = Tr\left(\hat{W}\hat{B}\right) = Tr_{+}\left(\hat{W}_{++}\hat{B}_{++}\right) + Tr_{-}\left(\hat{W}_{--}\hat{B}_{--}\right)$$

where $Tr_{\pm}$ implies a trace only over the particular subspace.

The cross matrix elements $\hat{W}_{+-}$ and $\hat{W}_{-+}$ do not contribute to the expectation value of the observables, or interference between vectors of the $|+\rangle$ and $|-\rangle$ types is not observable.

All equations of motion decouple into two separate equations in each of the two subspaces and no cross-matrix elements of $\hat{W}$ between the two subspaces ever contribute.

If we assume that the cross matrix elements are zero initially, then they will never develop(become nonzero) in time.

What is the difference between a generator $\hat{U}(2\pi)$ of a superselection rule and a symmetry operation that is generated by a universally conserved quantity such as the displacement operator

$$e^{-\frac{i}{\hbar}\vec{a}\cdot\vec{P}_{op}}$$

which is generated by the total momentum $\vec{P}_{op}$?

The Hamiltonian of any closed system is invariant under both transformations.

Both give rise to a quantum number that must be conserved in any transition.

In these examples the quantum numbers are

± 1 for $U(2\pi)$ and the total momentum

The difference is that there exist observables that do not commute with $\vec{P}_{op}$ and $\vec{Q}_{op}$, but there are no observables that do not commute with $\hat{U}(2\pi)$.

By measuring the position one can distinguish states that differ only by a displacement, but there is no way to distinguish between states that differ only by a 2π rotation.

The superselection rules from $\hat{U}(2\pi)$, which separates the integer and half-integer angular momentum states, is the only such rule in the quantum mechanics of stable particles (non-relativistic quantum mechanics).

In Quantum Field Theory(relativistic quantum mechanics), where particles can be created and annihilated, the total electric charge operator generates another superselection rule, provided that one assumes all observables are invariant under gauge transformations.

This says that no interference can be observed between states of different total charge because there are no observables that do not commute with the charge operator.

In a theory of stable particles, the charge of each particle and hence the total charge is an invariant.

Thus, the total charge operator is simply a multiple of $\hat{I}$.

Every operator commutes with $\hat{I}$ implying that the charge superselection rule is trivial in non-relativistic quantum mechanics.

Now back to the normal world.

The techniques that we have developed for the spin = 1/2 system can be applied to any two-state system.

Here is an example of a two-sided box solved using both the Schrodinger and Heisenberg pictures.

A Box with 2 Sides

Let us consider a box containing a particle in a state $|\psi\rangle$.

The box is divided into two halves (**R**ight and **L**eft) by a thin partition.

The only property that we will assign to the particle is whether it is on the **R**ight or **L**eft side of the box.

This means that the system has only two states, namely, $|R\rangle$ and $|L\rangle$ that must have the properties

$$\langle R \mid \psi \rangle = \text{ amplitude to find the particle on right side if in state } |\psi\rangle$$
$$\langle L \mid \psi \rangle = \text{ amplitude to find the particle on leftt side if in state } |\psi\rangle$$

We also suppose that the particle can tunnel through the thin partition and that

$$\frac{d}{dt}\langle R \mid \psi \rangle = \frac{K}{i\hbar}\langle L \mid \psi \rangle \quad , \quad K \text{ real}$$

How does this system develop in time?

We solve this problem in two ways, namely, using the Schrodinger and Heisenberg pictures.

Schrodinger Picture

We define the general system state vector

$$|\psi\rangle = \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} = \begin{pmatrix} \langle R \mid \psi \rangle \\ \langle L \mid \psi \rangle \end{pmatrix}$$
$$= \psi_R \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \psi_L \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$= \psi_R |R\rangle + \psi_L |L\rangle$$

The time-dependent Schrodinger equation is

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = i\hbar\begin{pmatrix}\frac{\partial\psi_R}{\partial t}\\ \frac{\partial\psi_L}{\partial t}\end{pmatrix} = \hat{H}\begin{pmatrix}\psi_R\\ \psi_L\end{pmatrix}$$

Now we are given that

$$\frac{d\psi_R}{dt} = \frac{K}{i\hbar}\psi_L \rightarrow \frac{d\psi_R^*}{dt} = -\frac{K}{i\hbar}\psi_L^*$$

The state vector must remain normalized as it develops in time so that

$$\langle\psi \mid \psi\rangle = |\psi_R|^2 + |\psi_L|^2 = 1$$

$$\frac{\partial\langle\psi \mid \psi\rangle}{\partial t} = 0 = \frac{\partial\psi_R^*}{\partial t}\psi_R + \psi_R^*\frac{\partial\psi_R}{\partial t} + \frac{\partial\psi_L^*}{\partial t}\psi_L + \psi_L^*\frac{\partial\psi_L}{\partial t}$$

$$0 = -\frac{K}{i\hbar}\psi_L^*\psi_R + \psi_R^*\frac{K}{i\hbar}\psi_L + \frac{\partial\psi_L^*}{\partial t}\psi_L + \psi_L^*\frac{\partial\psi_L}{\partial t}$$

$$0 = \psi_L\left(\frac{\partial\psi_L^*}{\partial t} + \frac{K}{i\hbar}\psi_R^*\right) + \psi_L^*\left(\frac{\partial\psi_L}{\partial t} - \frac{K}{i\hbar}\psi_R\right)$$

which says that

$$\frac{d\psi_L}{dt} = \frac{K}{i\hbar}\psi_R \rightarrow \frac{d\psi_L^*}{dt} = -\frac{K}{i\hbar}\psi_R^*$$

Therefore we have

$$K\begin{pmatrix}\psi_L\\ \psi_R\end{pmatrix} = \hat{H}\begin{pmatrix}\psi_R\\ \psi_L\end{pmatrix}$$

which says that

$$\hat{H} = K\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$$

The eigenvalues of $\hat{H}$ are $\pm K$ and its eigenvectors are

$$|\pm K\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}$$

Note that for $\hat{\wp}$ = parity operator (switches right and left) we have

$$\hat{\wp}\,|\pm K\rangle = \pm\,|\pm K\rangle$$

so these are also states of definite parity.

If the initial state of the system is $|\psi(0)\rangle = \begin{pmatrix} \psi_R(0) \\ \psi_L(0) \end{pmatrix}$

then we can write this state in terms of energy eigenstates as

$$|\psi(0)\rangle = \begin{pmatrix} \psi_R(0) \\ \psi_L(0) \end{pmatrix}$$

$$= \frac{1}{\sqrt{2}} \left(\psi_R(0) + \psi_L(0)\right) |+K\rangle + \frac{1}{\sqrt{2}} \left(\psi_R(0) - \psi_L(0)\right) |-K\rangle$$

Since we know the time dependence of energy eigenstates

$$|\pm K\rangle_t = e^{\mp \frac{i}{\hbar} Kt}\,|\pm K\rangle$$

the time dependence of $|\psi(t)\rangle$ is given by

$$|\psi(t)\rangle = \frac{1}{\sqrt{2}} \left(\psi_R(0) + \psi_L(0)\right) e^{-\frac{i}{\hbar} Kt}\,|+K\rangle + \frac{1}{\sqrt{2}} \left(\psi_R(0) - \psi_L(0)\right) e^{+\frac{i}{\hbar} Kt}\,|-K\rangle$$

or

$$|\psi(t)\rangle = \frac{1}{2}\begin{pmatrix} (\psi_R(0)+\psi_L(0))\,e^{-\frac{i}{\hbar}Kt} + (\psi_R(0)-\psi_L(0))\,e^{+\frac{i}{\hbar}Kt} \\ (\psi_R(0)+\psi_L(0))\,e^{-\frac{i}{\hbar}Kt} - (\psi_R(0)-\psi_L(0))\,e^{+\frac{i}{\hbar}Kt} \end{pmatrix}$$

$$= \left(\psi_R(0)\cos\frac{Kt}{\hbar} - i\psi_L(0)\sin\frac{Kt}{\hbar}\right)|R\rangle$$

$$+ \left(-i\psi_R(0)\sin\frac{Kt}{\hbar} + \psi_L(0)\cos\frac{Kt}{\hbar}\right)|L\rangle$$

Therefore, the probability that the particle is on the **R**ight side at time t is

$$P_R = |\langle R \mid \psi(t)\rangle|^2 = \left|\psi_R(0)\cos\frac{Kt}{\hbar} - i\psi_L(0)\sin\frac{Kt}{\hbar}\right|^2$$

Now suppose that

$$\psi_R(0) = \frac{1}{\sqrt{2}} \text{ and } \psi_L(0) = e^{i\delta}\psi_R(0)$$

This says that the particle was equally probable to be on either side at $t = 0$, but that the amplitudes differed by a phase factor.

In this case, we get

$$P_R = \frac{1}{2}\left(1 + \sin\delta\sin\frac{2Kt}{\hbar}\right)$$

Heisenberg Picture

Let us define an operator $\hat{Q}$ such that for a state vector

$$|\psi\rangle = \psi_R\,|R\rangle + \psi_L\,|L\rangle$$

we have

$$\langle\psi|\,\hat{Q}\,|\psi\rangle = |\psi_R|^2 = \text{ probability that particle is on right side}$$

This says that

$$\langle\psi|\,\hat{Q}\,|\psi\rangle = |\psi_R|^2 = \langle\psi\mid R\rangle\langle R\mid\psi\rangle$$

or

$$\hat{Q} = |R\rangle\langle R| = \text{ pure state projection operator } = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \hat{I} + \hat{\sigma}_z$$

and that the expectation value of the pure state projection operator is equal to the probability of being in that state.

This agrees with our earlier discussions.

Now we have

$$\hat{H} = K\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = K\hat{\sigma}_x$$

Therefore

$$\hat{Q}(t) = e^{\frac{i}{\hbar}\hat{H}t}\hat{Q}e^{-\frac{i}{\hbar}\hat{H}t}$$

Now as we saw earlier

$$e^{\frac{i}{\hbar}\hat{H}t} = e^{\frac{i}{\hbar}K\hat{\sigma}_x t} = \cos\frac{Kt}{\hbar}\hat{I} + i\sin\frac{Kt}{\hbar}\hat{\sigma}_x$$

Thus,

$$\begin{aligned}
\hat{Q}(t) &= \left(\cos\frac{Kt}{\hbar}\hat{I} + i\sin\frac{Kt}{\hbar}\hat{\sigma}_x\right)\left(\hat{I} + \hat{\sigma}_z\right)\left(\cos\frac{Kt}{\hbar}\hat{I} - i\sin\frac{Kt}{\hbar}\hat{\sigma}_x\right) \\
&= \cos^2\frac{Kt}{\hbar} + \sin^2\frac{Kt}{\hbar} + \cos^2\frac{Kt}{\hbar}\hat{\sigma}_z - i\sin\frac{Kt}{\hbar}\cos\frac{Kt}{\hbar}\left[\hat{\sigma}_z, \hat{\sigma}_x\right] + \sin^2\frac{Kt}{\hbar}\hat{\sigma}_x\hat{\sigma}_z\hat{\sigma}_x \\
&= 1 + \cos^2\frac{Kt}{\hbar}\hat{\sigma}_z + 2\sin\frac{Kt}{\hbar}\cos\frac{Kt}{\hbar}\hat{\sigma}_y - \sin^2\frac{Kt}{\hbar}\hat{\sigma}_z \\
&= 1 + \cos\frac{2Kt}{\hbar}\hat{\sigma}_z + \sin\frac{2Kt}{\hbar}\hat{\sigma}_y
\end{aligned}$$

Then

$$P_R(t) = \langle\psi(0)|\,\hat{Q}(t)\,|\psi(0)\rangle$$

where

$$|\psi(0)\rangle = \frac{1}{\sqrt{2}}|R\rangle + \frac{e^{i\delta}}{\sqrt{2}}|L\rangle$$

Now

$$\hat{\sigma}_z|R\rangle = |R\rangle \quad , \quad \hat{\sigma}_z|L\rangle = -|L\rangle$$
$$\hat{\sigma}_y|R\rangle = -i|L\rangle \quad , \quad \hat{\sigma}_y|L\rangle = i|R\rangle$$

and we get

$$P_R = \frac{1}{2}(1 + \sin\delta \sin\frac{2Kt}{\hbar})$$

as before.

Magnetic Resonance

How does an experimentalist observe the spin of a particle?

Classically, a spinning charge distribution will have an associated magnetic moment.

In non-relativistic quantum mechanics particles with internal spin degrees of freedom also have magnetic moments which are connected to their angular momentum.

We write for the magnetic moment operator

$$\vec{M}_{op} = \vec{M}_{op}^{orbital} + \vec{M}_{op}^{spin}$$
$$= g_\ell \frac{q}{2mc}\vec{L}_{op} + g_s \frac{q}{2mc}\vec{S}_{op} = \frac{q}{2mc}\left(g_\ell \vec{L}_{op} + g_s \vec{S}_{op}\right)$$

where, as we shall derive later,

$$g_{j\ell s} = 1 + \frac{j(j+1) - \ell(\ell+1) + s(s+1)}{2j(j+1)}$$

$$g_\ell = g_{jj0} = 1 \text{ and } g_s = g_{j0j} = 2$$

Therefore we have

$$\vec{M}_{op} = \frac{q}{2mc}\left(\vec{L}_{op} + 2\vec{S}_{op}\right) = \frac{q}{2mc}\left(\vec{J}_{op} + \vec{S}_{op}\right)$$

which says that $\vec{M}_{op}$ is not parallel to $\vec{J}_{op}$.

The energy operator or contribution to the Hamiltonian operator from the magnetic moment in a magnetic field $\vec{B}(\vec{r},t)$ is

$$\hat{H}_M = -\vec{M}_{op} \cdot \vec{B}(\vec{r},t)$$

Spin measurement experiments are designed to detect the effects of this extra contribution to the system energy and thus detect the effects of spin angular momentum.

We will return to a full discussion of the effect of $\hat{H}_M$ on the energy levels of atoms, etc later.

For now we will restrict our attention to spin space and investigate the effect of the spin contribution to $\hat{H}_M$ on the states of a particle.

We will use

$$\hat{H}_{spin} = -\frac{g}{2}\frac{q}{mc}\vec{B} \cdot \bar{S}_{op}$$

where we have replaced g_s by g.

We have not set $g = 2$ since it turns out that it is not exactly that value(due to relativistic effects).

In the Schrodinger picture we have

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H}_{spin} |\psi(t)\rangle$$

If we ignore spatial dependences (worry only about spin effects), we have

$$|\psi(t)\rangle = \begin{pmatrix} \langle \hat{n}+ \mid \psi(t)\rangle \\ \langle \hat{n}- \mid \psi(t)\rangle \end{pmatrix}$$

$$i\hbar \frac{d}{dt} \begin{pmatrix} \langle \hat{n}+ \mid \psi(t)\rangle \\ \langle \hat{n}- \mid \psi(t)\rangle \end{pmatrix} = -\frac{g}{2}\frac{q}{mc} \vec{B} \cdot \bar{S}_{op} \begin{pmatrix} \langle \hat{n}+ \mid \psi(t)\rangle \\ \langle \hat{n}- \mid \psi(t)\rangle \end{pmatrix}$$

$$= -\frac{g}{4}\frac{q\hbar}{mc} \vec{B} \cdot \vec{\sigma}_{op} \begin{pmatrix} \langle \hat{n}+ \mid \psi(t)\rangle \\ \langle \hat{n}- \mid \psi(t)\rangle \end{pmatrix}$$

$$= -\frac{g}{4}\frac{q\hbar}{mc} \begin{pmatrix} B_z & B_x - iB_y \\ B_x + iB_y & -B_z \end{pmatrix} \begin{pmatrix} \langle \hat{n}+ \mid \psi(t)\rangle \\ \langle \hat{n}- \mid \psi(t)\rangle \end{pmatrix}$$

This represents two coupled differential equations for the time dependence of the amplitudes $\langle + | \psi \rangle$ and $\langle - | \psi \rangle$. When solved, the solution tells us the time dependence of the measurable probabilities.

We can see the physics of the motion best in the Heisenberg picture, where the operators, instead of the states, move in time.

In this case, we have

$$i\hbar \frac{d\hat{S}_i(t)}{dt} = \left[\hat{S}_i(t), \hat{H}_{spin}(t)\right]$$

$$= -\frac{gq}{2mc} \left[\hat{S}_i(t), \hat{S}_j(t)\right] B_j(t) = -i\frac{gq\hbar}{2mc} \varepsilon_{ijk} \hat{S}_k(t) B_j(t)$$

for each component.

The operators are all time-dependent Heisenberg picture operators.

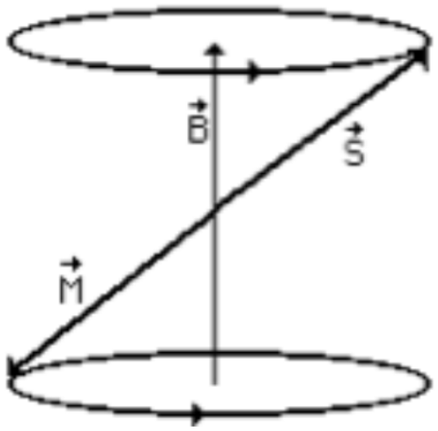
This gives

$$\frac{d\vec{S}_{op}(t)}{dt} = \frac{gq}{2mc}\vec{S}_{op}(t) \times \vec{B}(t) = \vec{M}_{spin}(t) \times \vec{B}(t)$$

The right-hand-side is the *torque* exerted by the magnetic field on the magnetic moment.

In operator language, this equation implies that the rate of change of the spin angular momentum vector equals the applied torque.

This implies that the spin vector with $q < 0$ precesses in a positive sense about the magnetic field direction as shown in the Figure below.



Motion of Spin

Suppose that $\vec{B}(t) = B_0\ \hat{z}$ (independent of t). We then have

$$\frac{d\hat{S}_z(t)}{dt} = 0$$

$$\frac{d\hat{S}_x(t)}{dt} = \frac{gqB_0}{2mc}\hat{S}_y(t)$$

$$\frac{d\hat{S}_y(t)}{dt} = -\frac{gqB_0}{2mc}\hat{S}_x(t)$$

These coupled differential equations have the solutions

$$\hat{S}_x(t) = \hat{S}_x(0)\cos\omega_0 t + \hat{S}_y(0)\sin\omega_0 t$$
$$\hat{S}_y(t) = -\hat{S}_x(0)\sin\omega_0 t + \hat{S}_y(0)\cos\omega_0 t$$
$$\hat{S}_z(t) = \hat{S}_z(0)$$

where

$$\omega_0 = \frac{gqB_0}{2mc}$$

Now suppose that at $t = 0$, the spin is in the $+x$-direction, which says that

$$\left\langle \hat{S}_x(0) \right\rangle = \langle \hat{x}+ | \hat{S}_x(0) | \hat{x}+ \rangle = \frac{\hbar}{2}$$
$$\left\langle \hat{S}_y(0) \right\rangle = \left\langle \hat{S}_z(0) \right\rangle = 0$$

Therefore, the expectation values of the solutions become

$$\left\langle \hat{S}_x(t) \right\rangle = \frac{\hbar}{2}\cos\omega_0 t$$
$$\left\langle \hat{S}_y(t) \right\rangle = -\frac{\hbar}{2}\sin\omega_0 t$$
$$\left\langle \hat{S}_z(0) \right\rangle = 0$$

which says that the expectation value of the spin vector rotates in a negative sense in the x–y plane (precession about the z-axis).

Now let us return to the Schrodinger picture to see what precession looks like there.

We will do the calculation a couple of different ways.

First we use the time-development operator. For $\vec{B} = B\,\hat{n}$ we have

$$\hat{H} = -\frac{gq\hbar B}{4mc}\vec{\sigma}_{op}\cdot\hat{n}$$

If we define

$$\omega_L = \frac{qB}{mc} = \text{ Larmor frequency}$$

and let $g = 2$ we have

$$\hat{H} = -\frac{1}{2}\hbar\omega_L\vec{\sigma}_{op}\cdot\hat{n}$$

In the Schrodinger picture

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = \hat{H}_{spin}(t)\,|\psi(t)\rangle$$

When $\hat{H}_{\text{spin}}(t)$ is time-independent, we have the solution

$$|\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle$$

where

$$\hat{U}(t) = e^{-\frac{i}{\hbar}\hat{H}t}$$

or

$$|\psi(t)\rangle = e^{i\frac{\omega_L t}{2}\vec{\sigma}_{op}\cdot\hat{n}}\,|\psi(0)\rangle$$

Since $\hat{n}$ = a unit vector, we have $(\vec{\sigma}_{\text{op}}\cdot\hat{n}) = \hat{I}$, which gives, as before,

$$|\psi(t)\rangle = \left(\cos\frac{\omega_L t}{2} + i(\vec{\sigma}_{op}\cdot\hat{n})\sin\frac{\omega_L t}{2}\right)|\psi(0)\rangle$$

This is the solution to the Schrodinger equation in this case.

Now let

$$|\psi(0)\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |\hat{z}\uparrow\rangle = |\hat{z}+\rangle = |+\rangle$$

From earlier we have

$$\vec{\sigma}_{op}\cdot\hat{n} = \begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix}$$

We then have

$$\begin{aligned} |\psi(t)\rangle &= \cos\frac{\omega_L t}{2}|+\rangle + i\begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix}|+\rangle \sin\frac{\omega_L t}{2} \\ &= \cos\frac{\omega_L t}{2}|+\rangle + i\begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\sin\frac{\omega_L t}{2} \\ &= \cos\frac{\omega_L t}{2}|+\rangle + i\sin\frac{\omega_L t}{2}\begin{pmatrix} n_z \\ n_x + in_y \end{pmatrix} \\ &= \cos\frac{\omega_L t}{2}|+\rangle + i\sin\frac{\omega_L t}{2}\left(n_z|+\rangle + (n_x + in_y)|-\rangle\right) \\ &= \left(\cos\frac{\omega_L t}{2} + in_z\sin\frac{\omega_L t}{2}\right)|+\rangle + i(n_x + in_y)\sin\frac{\omega_L t}{2}|-\rangle \end{aligned}$$

This says that as the initial state $|+\rangle$ develops in time it picks up some amplitude to be in the $|-\rangle$ state.

Special Case

Let

$$\hat{n} = \hat{y} \rightarrow \vec{B} \text{ is in the } y\text{- direction}$$

$$n_y = 1\ ,\ n_x = n_z = 0$$

This gives

$$|\psi(t)\rangle = \cos\frac{\omega_L t}{2}|+\rangle - \sin\frac{\omega_L t}{2}|-\rangle$$

which implies that $\langle \hat{s}_z \rangle$ flips with frequency $\nu = \omega_L/4\pi$ or that the spin vector is precessing around the direction of the magnetic field.

Finally, let us just solve the Schrodinger equation directly.

We have

$$i\hbar\frac{d}{dt}\begin{pmatrix} \langle \hat{n}+ \mid \psi(t)\rangle \\ \langle \hat{n}- \mid \psi(t)\rangle \end{pmatrix} = -\frac{1}{2}\frac{q\hbar}{mc}\begin{pmatrix} B_z & B_x - iB_y \\ B_x + iB_y & -B_z \end{pmatrix}\begin{pmatrix} \langle \hat{n}+ \mid \psi(t)\rangle \\ \langle \hat{n}- \mid \psi(t)\rangle \end{pmatrix}$$

If we choose $\hat{n} = \hat{x}$, then we have

$$i\hbar\frac{d}{dt}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = -\frac{1}{2}\hbar\omega_L\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix}$$

where $a(t) = \langle \hat{n}+ \mid \psi(t)\rangle$ and $b(t) = \langle \hat{n}- \mid \psi(t)\rangle$

and $|a|^2 + |b|^2 = 1$

This gives two coupled differential equations

$$\dot{a} = i\frac{\omega_L}{2}b \text{ and } \dot{b} = i\frac{\omega_L}{2}a$$

which are solved as follows:

$$(\dot{a} + \dot{b}) = i\frac{\omega_L}{2}(\dot{a} + \dot{b}) \rightarrow a + b = (a(0) + b(0))e^{i\frac{\omega_L}{2}t}$$

$$(\dot{a} - \dot{b}) = -i\frac{\omega_L}{2}(\dot{a} - \dot{b}) \rightarrow a - b = (a(0) - b(0))e^{-i\frac{\omega_L}{2}t}$$

or

$$a(t) = a(0)\cos\frac{\omega_L}{2}t + ib(0)\sin\frac{\omega_L}{2}t$$

$$b(t) = ia(0)\sin\frac{\omega_L}{2}t + b(0)\cos\frac{\omega_L}{2}t$$

For the initial state

$$|\psi(0)\rangle = \begin{pmatrix} a(0) \\ b(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

we get

$$a(t) = \cos\frac{\omega_L}{2}t \quad , \quad b(t) = i\sin\frac{\omega_L}{2}t$$

and

$$a(t) = \cos\frac{\omega_L}{2}t \quad , \quad b(t) = i\sin\frac{\omega_L}{2}t$$

as before.

Spin Resonance

Let us now consider the addition of an oscillating magnetic field perpendicular to the applied field $\vec{B} = B_0\,\hat{z}$.

In particular, we add the field

$$\vec{B}_1 = B_1\cos\omega t\hat{x}$$

In the Schrodinger picture, we then have the equation of motion

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = -\frac{e\hbar}{2mc}\frac{g}{2}\left(B_0\hat{\sigma}_z + B_1\cos\omega t\hat{\sigma}_x\right)|\psi(t)\rangle$$

Now we first make the transformation (shift to rotating coordinate system since we know that if the extra field were not present the spin would just be precessing about the direction of the applied field)

$$|\psi(t)\rangle = e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle$$

which gives

$$i\hbar\frac{d}{dt}\left(e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle\right) = -\frac{e\hbar}{2mc}\frac{g}{2}\left(B_0\hat{\sigma}_z + B_1\cos\omega t\hat{\sigma}_x\right)e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle$$

$$i\hbar e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\frac{d}{dt}\left(e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle\right)$$

$$= -\frac{e\hbar}{2mc}\frac{g}{2}\left(B_0 e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\hat{\sigma}_z e^{i\frac{\omega t}{2}\hat{\sigma}_z} + B_1\cos\omega t e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\hat{\sigma}_x e^{i\frac{\omega t}{2}\hat{\sigma}_z}\right)|\psi'(t)\rangle$$

or

$$i\hbar e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\left(e^{i\frac{\omega t}{2}\hat{\sigma}_z}\frac{d}{dt}|\psi'(t)\rangle + i\frac{\omega}{2}\hat{\sigma}_z e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle\right)$$

$$= -\frac{e\hbar}{2mc}\frac{g}{2}\left(B_0\hat{\sigma}_z + B_1\cos\omega t e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\hat{\sigma}_x e^{i\frac{\omega t}{2}\hat{\sigma}_z}\right)|\psi'(t)\rangle$$

Now

$$\begin{aligned}
\cos\omega t e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\hat{\sigma}_x e^{i\frac{\omega t}{2}t\hat{\sigma}_z} &= \cos\omega t\hat{\sigma}_x e^{i\omega t\hat{\sigma}_z} \\
&= \hat{\sigma}_x(\cos^2\omega t + i\hat{\sigma}_z\cos\omega t\sin\omega t) \\
&= \hat{\sigma}_x(\frac{1}{2} + \frac{1}{2}\cos 2\omega t + i\hat{\sigma}_z\frac{1}{2}\sin 2\omega t) \\
&= \frac{\hat{\sigma}_x}{2} + \frac{1}{2}(\hat{\sigma}_x\cos 2\omega t + i\hat{\sigma}_x\hat{\sigma}_z\sin 2\omega t) \\
&= \frac{\hat{\sigma}_x}{2} + \frac{1}{2}(\hat{\sigma}_x\cos 2\omega t + \hat{\sigma}_y\sin 2\omega t)
\end{aligned}$$

Defining

$$\omega_0 = \frac{geB_0}{4mc} \text{ and } \omega_1 = \frac{geB_1}{4mc}$$

we get

$$i\hbar \frac{d}{dt} |\psi'(t)\rangle$$
$$= \left(\left(\frac{\omega - \omega_0}{2} \right) \hat{\sigma}_z - \frac{\omega_1}{2} \hat{\sigma}_x \right) |\psi'(t)\rangle$$
$$+ \frac{\omega_1}{2} \left(\hat{\sigma}_x \cos 2\omega t + \hat{\sigma}_y \sin 2\omega t \right) |\psi'(t)\rangle$$

The two higher frequency terms produce high frequency wiggles in $|\psi'(t)\rangle$.

Since we will be looking at the average motion of the spin vector (expectation values changing in time), these terms will average to zero and we can neglect them in this discussion.

We thus have

$$i\hbar \frac{d}{dt} |\psi'(t)\rangle = \left(\left(\frac{\omega - \omega_0}{2} \right) \hat{\sigma}_z - \frac{\omega_1}{2} \hat{\sigma}_x \right) |\psi'(t)\rangle$$

which has a solution

$$|\psi'(t)\rangle = e^{-i\frac{\Omega t}{2}\hat{\sigma}} |\psi'(0)\rangle$$

$$\Omega = \left[(\omega - \omega_0)^2 + \omega_1^2 \right]^{1/2} \text{ and } \hat{\sigma} = \frac{\omega - \omega_0}{\Omega} \hat{\sigma}_z - \frac{\omega_1}{\Omega} \hat{\sigma}_x$$

We note that

$$\hat{\sigma}^2 = \frac{(\omega - \omega_0)^2 + \omega_1^2}{\Omega^2} \hat{I} = \hat{I}$$

The final solution to the original problem (again neglecting the higher frequency terms) is then(after leaving the rotating system)

$$|\psi(t)\rangle = e^{-i\frac{\omega t}{2}\hat{\sigma}_z} e^{-i\frac{\Omega t}{2}\hat{\sigma}} |\psi(0)\rangle$$

Example

Let us choose

$$|\psi(0)\rangle = |+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

We get

$$|\psi(t)\rangle = e^{-i\frac{\omega t}{2}\hat{\sigma}_z} e^{-i\frac{\Omega t}{2}\hat{\sigma}} |+\rangle = e^{-i\frac{\omega t}{2}\hat{\sigma}_z} \left(\cos\frac{\Omega t}{2} - i\hat{\sigma}\sin\frac{\Omega t}{2}\right)|+\rangle$$

$$= e^{-i\frac{\omega t}{2}\hat{\sigma}_z} \left(\cos\frac{\Omega t}{2}|+\rangle - i\frac{\omega-\omega_0}{\Omega}\sin\frac{\Omega t}{2}\hat{\sigma}_z|+\rangle + i\frac{\omega_1}{\Omega}\sin\frac{\Omega t}{2}\hat{\sigma}_x|+\rangle\right)$$

Now using

$$\hat{\sigma}_z|+\rangle = |+\rangle \text{ and } \hat{\sigma}_x|+\rangle = |-\rangle$$

$$e^{-i\frac{\omega t}{2}\hat{\sigma}_z} = \cos\frac{\omega t}{2} - i\sin\frac{\omega t}{2}\hat{\sigma}_z$$

we have

$$|\psi(t)\rangle = A^{(+)}|+\rangle + A^{(-)}|-\rangle$$

$$A^{(+)} = \cos\frac{\omega t}{2}\cos\frac{\Omega t}{2} - i\sin\frac{\omega t}{2}\cos\frac{\Omega t}{2}$$

$$- i\frac{\omega-\omega_0}{\Omega}\cos\frac{\omega t}{2}\sin\frac{\Omega t}{2} - \frac{\omega-\omega_0}{\Omega}\sin\frac{\omega t}{2}\sin\frac{\Omega t}{2}$$

$$A^{(-)} = -i\frac{\omega_1}{\Omega}\cos\frac{\omega t}{2}\sin\frac{\Omega t}{2} + \frac{\omega_1}{\Omega}\sin\frac{\omega t}{2}\sin\frac{\Omega t}{2}$$

Therefore, the amplitude for spin flip to the state $|\hat{z}\downarrow\rangle = |-\rangle$ at time t is

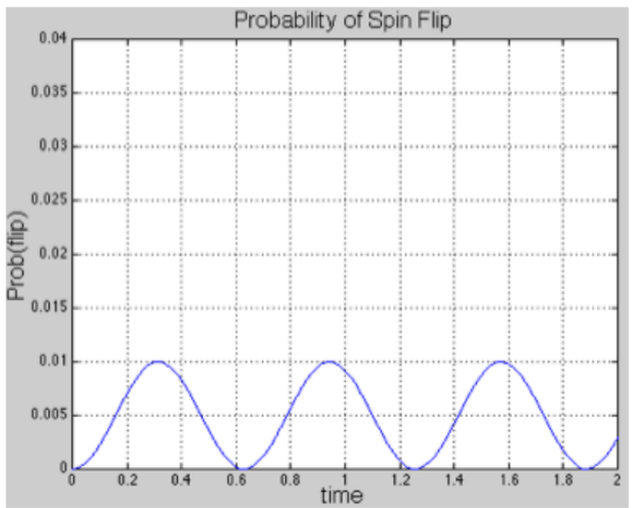
$$\langle - \,|\, \psi(t)\rangle = A^{(-)} = -i\frac{\omega_1}{\Omega}\cos\frac{\omega t}{2}\sin\frac{\Omega t}{2} + \frac{\omega_1}{\Omega}\sin\frac{\omega t}{2}\sin\frac{\Omega t}{2}$$

$$= -i\frac{\omega_1}{\Omega}\sin\frac{\Omega t}{2}e^{i\frac{\omega t}{2}}$$

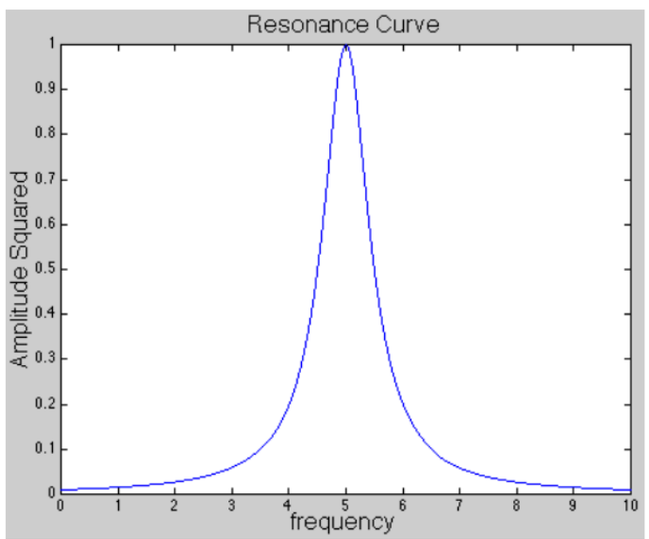
and the probability of spin flip is $$P_{flip}(t) = |\langle - \,|\, \psi(t)\rangle|^2 = \frac{\omega_1^2}{\Omega^2}\sin^2\frac{\Omega t}{2} = \frac{\omega_1^2}{2\Omega^2}(1-\cos\Omega t)$$

What is happening to the spin vector?

If we plot $P_{\text{flip}}(t)$ versus t as in the Figure below left we get



Spin=Flip Probability versus Time



Resonance Curve

where the peak values are given by ω_1^2/Ω^2.

What is the value of ω_1^2/Ω^2 and can it be large? This spin resonance process is used to determine a wide variety of spin properties of systems.

If we plot ω_1^2/Ω^2 versus ω (the frequency of the added field) we get as in Figure above right the result The peak occurs at $\omega - \omega_0 \approx 0$ (called a *resonance*).

Therefore, if $\omega - \omega_0 \gg \omega_1$ (called off-resonance), the maximum probability for spin flip is small(that corresponds to the Figure left).

However, if $\omega - \omega_0 \approx 0$, which corresponds to resonance, the maximum probability ≈ 1 and the spin has flipped with certainty.

The spin system preferentially absorbs energy (flipping spin) near resonance.
This spin resonance process is used to determine a wide variety of spin properties of systems.

Addition of Angular Momentum

A derivation of the addition process for arbitrary angular momentum values is very complex.

We can, however, learn and understand all of the required steps within the context of a special case, namely, combining two spin = 1/2 systems into a new system.

We will do the general derivation after the special case.

Addition of Two Spin = 1/2 Angular Momenta

We define

$$\vec{S}_{1,op} = \text{ spin operator for system 1}$$
$$\vec{S}_{2,op} = \text{ spin operator for system 2}$$

The operators for system 1 are assumed to be independent of the operators for system 2, which implies that

$$\left[\hat{S}_{1i}, \hat{S}_{2j}\right] = 0 \text{ for all } i \text{ and } j$$

We characterize each system by eigenvalue/eigenvector equations

$$\vec{S}^2_{1,op}\left|s_1, m_1\right\rangle = \hbar^2 s_1(s_1+1)\left|s_1, m_1\right\rangle$$
$$\hat{S}_{1,z}\left|s_1, m_1\right\rangle = \hbar m_1 \left|s_1, m_1\right\rangle$$
$$\vec{S}^2_{2,op}\left|s_2, m_2\right\rangle = \hbar^2 s_2(s_2+1)\left|s_2, m_2\right\rangle$$
$$\hat{S}_{2,z}\left|s_2, m_2\right\rangle = \hbar m_2 \left|s_2, m_2\right\rangle$$

where

$$s_1 = s_2 = \frac{1}{2} \text{ and } m_1 = \pm\frac{1}{2}\,,\; m_2 = \pm\frac{1}{2}$$

Since both s_1 and s_2 are fixed and unchanging during this addition process we will drop them from the arguments and subscripts to lessen the complexity of the equations.

Each space (1 and 2) is 2-dimensional and thus each has two ($2s + 1 = 2$) basis states corresponding to the number of m-values in each case

$$\left|s_1 = \frac{1}{2}, m_1 = \frac{1}{2}\right\rangle = \left|\frac{1}{2}, \frac{1}{2}\right\rangle_1 = |\uparrow\rangle_1 = |+\rangle_1$$
$$\left|s_1 = \frac{1}{2}, m_1 = -\frac{1}{2}\right\rangle = \left|\frac{1}{2}, -\frac{1}{2}\right\rangle_1 = |\downarrow\rangle_1 = |-\rangle_1$$
$$\left|s_2 = \frac{1}{2}, m_2 = \frac{1}{2}\right\rangle = \left|\frac{1}{2}, \frac{1}{2}\right\rangle_2 = |\uparrow\rangle_2 = |+\rangle_2$$
$$\left|s_2 = \frac{1}{2}, m_2 = -\frac{1}{2}\right\rangle = \left|\frac{1}{2}, -\frac{1}{2}\right\rangle_2 = |\downarrow\rangle_2 = |-\rangle_2$$

This means that there are 4 possible basis states for the combined system that we can construct using the direct product procedure.

We label them as

$$|\uparrow\uparrow\rangle = |++\rangle = |+\rangle_1 \otimes |+\rangle_2 \quad , \quad |\uparrow\downarrow\rangle = |+-\rangle = |+\rangle_1 \otimes |-\rangle_2$$
$$|\downarrow\uparrow\rangle = |-+\rangle = |-\rangle_1 \otimes |+\rangle_2 \quad , \quad |\downarrow\downarrow\rangle = |--\rangle = |-\rangle_1 \otimes |-\rangle_2$$

so that the first symbol in the combined states corresponds to system 1 and the second symbol to system 2.

These are not the only possible basis states as we shall see.

The "1" operators operate only on the "1" part of the direct product, for example,

$$\vec{S}_{1,op}^2 |+-\rangle = \hbar^2 s_1(s_1+1) |+-\rangle = \frac{3}{4}\hbar^2 |+-\rangle$$

$$\hat{S}_{1,z} |+-\rangle = \hbar m_1 |+-\rangle = \frac{\hbar}{2} |+-\rangle$$

$$\vec{S}_{2,op}^2 |+-\rangle = \hbar^2 s_2(s_2+1) |+-\rangle = \frac{3}{4}\hbar^2 |+-\rangle$$

$$\hat{S}_{2,z} |+-\rangle = \hbar m_2 |+-\rangle = -\frac{\hbar}{2} |+-\rangle$$

The total spin angular momentum operator for the combined system is

$$\vec{S}_{op} = \vec{S}_{1,op} + \vec{S}_{2,op}$$

It obeys the same commutation rules as the individual system operators, i.e.,

$$\left[\hat{S}_i, \hat{S}_j\right] = i\hbar\varepsilon_{ijk}\hat{S}_k$$

This tells us to look for the same kind of angular momentum eigenstates and eigenvalues

$$\vec{S}_{op}^2 |s,m\rangle = \hbar^2 s(s+1) |s,m\rangle$$

$$\hat{S}_z |s,m\rangle = \hbar m |s,m\rangle$$

To proceed, we need to derive some relations.

Squaring the total spin operator we have

$$\vec{S}_{op}^2 = \left(\vec{S}_{1,op} + \vec{S}_{2,op}\right)^2 = \vec{S}_{1,op}^2 + \vec{S}_{2,op}^2 + 2\vec{S}_{1,op} \cdot \vec{S}_{2,op}$$

$$= \frac{3}{4}\hbar^2 \hat{I} + \frac{3}{4}\hbar^2 \hat{I} + 2\hat{S}_{1z}\hat{S}_{2z} + 2\hat{S}_{1x}\hat{S}_{2x} + 2\hat{S}_{1y}\hat{S}_{2y}$$

Now using

$$\hat{S}_{1\pm} = \hat{S}_{1x} \pm i\hat{S}_{1y} \text{ and } \hat{S}_{2\pm} = \hat{S}_{2x} \pm i\hat{S}_{2y}$$

we have

$$2\hat{S}_{1x}\hat{S}_{2x} + 2\hat{S}_{1y}\hat{S}_{2y} = \hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+}$$

and therefore

$$\vec{S}_{op}^{2} = \frac{3}{2}\hbar^2\hat{I} + 2\hat{S}_{1z}\hat{S}_{2z} + \hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+}$$

Now suppose we choose the four states $|{+}{+}\rangle$, $|{+}{-}\rangle$, $|{-}{+}\rangle$, $|{-}{-}\rangle$ as the orthonormal basis for the 4-dimensional vector space of the combined system.

We then ask the following question.

Are these basis states also eigenstates of the spin operators for the combined system?

We have

$$\begin{aligned}\hat{S}_z|{+}{+}\rangle &= (\hat{S}_{1z} + \hat{S}_{2z})|{+}\rangle_1 \otimes |{+}\rangle_2 = (\hat{S}_{1z}|{+}\rangle_1) \otimes |{+}\rangle_2 + |{+}\rangle_1 \otimes (\hat{S}_{2z}|{+}\rangle_2) \\ &= (\frac{\hbar}{2}|{+}\rangle_1) \otimes |{+}\rangle_2 + |{+}\rangle_1 \otimes (\frac{\hbar}{2}|{+}\rangle_2) = (\frac{\hbar}{2} + \frac{\hbar}{2})|{+}\rangle_1 \otimes |{+}\rangle_2 \\ &= \hbar|{+}{+}\rangle\end{aligned}$$

and similarly

$$\hat{S}_z|{+}{-}\rangle = 0$$

$$\hat{S}_z|{-}{+}\rangle = 0$$

$$\hat{S}_z|{-}{-}\rangle = -\hbar|{-}{-}\rangle$$

This says that the direct-product basis states are eigenvectors of (total) $\hat{S}_z$ and that the eigenvalues are $m = 1, 0, -1$ and 0 (a second time).

Since each value s of the total angular momentum of the combined system must have $2s + 1$ associated m-values, these results tell us that the combined system will have

$$s = 1 \rightarrow m = 1, 0, -1 \text{ and } s = 0 \rightarrow m = 0$$

which accounts for all of the four states.

What about the $\vec{S}^2_{op}$ operator?

We have

$$\vec{S}^2_{op} |{+}{+}\rangle = \left(\frac{3}{2}\hbar^2 \hat{I} + 2\hat{S}_{1z}\hat{S}_{2z} + \hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+} \right) |{+}{+}\rangle$$

$$= \left(\frac{3}{2}\hbar^2 + 2\frac{\hbar}{2}\frac{\hbar}{2} + \hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+} \right) |{+}{+}\rangle$$

Now using

$$\hat{S}_{1+} |{+}{+}\rangle = 0 = \hat{S}_{+} |{+}{+}\rangle$$

we get

$$\vec{S}^2_{op} |{+}{+}\rangle = 2\hbar^2 |{+}{+}\rangle = \hbar^2 1(1+1) |{+}{+}\rangle$$

or the state $|{+}{+}\rangle$ is also an eigenstate of $\vec{S}^2_{op}$ with $s = 1$

$$|{+}{+}\rangle = |s = 1, m = 1\rangle = |1, 1\rangle$$

Similarly, we have

$$|{-}{-}\rangle = |s = 1, m = -1\rangle = |1, -1\rangle$$

or the state $|{-}{-}\rangle$ is also an eigenstate of $\vec{S}^2_{op}$ with $s = 1$

In the same manner we can show that $|{+}{-}\rangle$ and $|{-}{+}\rangle$ are not eigenstates of $\vec{S}^2_{op}$.

So the simple direct-product states are *not* appropriate to describe the combined system if we want to characterize it using

$$\vec{S}^2_{op} \text{ and } \hat{S}_z$$

However, since the direct-product states are a complete basis set, we should be able to construct the remaining two eigenstates of $\vec{S}^2_{op}$ and $\hat{S}_z$ as linear combinations of the direct-product states

$$|s,m\rangle = \sum_{m_1,m_2} a_{smm_1m_2} |m_1,m_2\rangle$$

where we have left out the s_1 and s_2 dependence in the states and coefficients.

In a formal manner, we can identify the so-called *Clebsch-Gordon coefficients* $a_{smm_1m_2}$ by using the orthonormality of the direct-product basis states.

We have

$$\langle m_1', m_2' \mid s,m\rangle = \sum_{m_1,m_2} a_{smm_1m_2} \langle m_1', m_2' \mid m_1,m_2\rangle$$

$$= \sum_{m_1,m_2} a_{smm_1m_2} \delta_{m_1'm_1} \delta_{m_2'm_2} = a_{smm_1'm_2'}$$

where we have used

$$\langle m_1', m_2' \mid m_1,m_2\rangle = \delta_{m_1'm_1} \delta_{m_2'm_2}$$

This does not help us actually compute the coefficients because we do not know the states $|s,m\rangle$.

A procedure that works in this case and that can be generalized, is the following.

We already found that $|{+}{+}\rangle = |1,1\rangle$ and $|{-}{-}\rangle = |1,-1\rangle$.

Now we define the operators

$$\hat{S}_\pm = \hat{S}_x \pm i\hat{S}_y = (\hat{S}_{1x} + \hat{S}_{2x}) \pm i(\hat{S}_{1y} + \hat{S}_{2y}) = \hat{S}_{1\pm} + \hat{S}_{2\pm}$$

These are the raising and lowering operators for the combined system and thus they satisfy the relations

$$\hat{S}_{\pm}|s,m\rangle = \hbar\sqrt{s(s+1)-m(m\pm 1)}\,|s,m\pm 1\rangle$$

This gives

$$\hat{S}_{+}|1,1\rangle = 0 = \hat{S}_{-}|1,-1\rangle$$

as we expect.

If, however, we apply $\hat{S}_{-}$ to the topmost state (maximum s and m values) we get

$$\begin{aligned}\hat{S}_{-}|1,1\rangle &= \hbar\sqrt{1(1+1)-1(1-1)}\,|1,0\rangle = \hbar\sqrt{2}\,|1,0\rangle \\ &= (\hat{S}_{1-}+\hat{S}_{2-})|+\rangle_1\otimes|+\rangle_2 = (\hat{S}_{1-}|+\rangle_1)\otimes|+\rangle_2 + |+\rangle_1\otimes(\hat{S}_{2-}|+\rangle_2) \\ &= \hbar\sqrt{\frac{1}{2}(\frac{1}{2}+1)-\frac{1}{2}(\frac{1}{2}-1)}\,|-\rangle_1\otimes|+\rangle_2 \\ &\quad + \hbar\sqrt{\frac{1}{2}(\frac{1}{2}+1)-\frac{1}{2}(\frac{1}{2}-1)}\,|+\rangle_1\otimes|-\rangle_2 \\ &= \hbar\,|-+\rangle + \hbar\,|+-\rangle\end{aligned}$$

or

$$|1,0\rangle = |s=1,m=0\rangle = \frac{1}{\sqrt{2}}|-+\rangle + \frac{1}{\sqrt{2}}|+-\rangle$$

Note that the only terms that appear on the right hand side are those that have $m = m_1 + m_2$.

We can easily see that this is a general property since

$$\begin{aligned}\hat{S}_z|s,m\rangle = m\hbar\,|s,m\rangle &= \sum_{m_1,m_2} a_{smm_1m_2}(\hat{S}_{1z}+\hat{S}_{2z})|m_1,m_2\rangle \\ &= \hbar\sum_{m_1,m_2} a_{smm_1m_2}(m_1+m_2)|m_1,m_2\rangle\end{aligned}$$

The only way to satisfy this equation is for $m = m_1 + m_2$ in every term in the sum.

Thus, our linear combination should really be written as a single sum of the form

$$|s,m\rangle = \sum_{\substack{m_1,m_2 \\ m_1+m_2=m}} a_{smm_1m_2} |m_1,m_2\rangle = \sum_{m_1} a_{smm_1,m-m_1} |m_1,m-m_1\rangle$$

These three states

$$|1,1\rangle = |{+}{+}\rangle \rightarrow a_{1,1,\frac{1}{2},\frac{1}{2}} = 1$$

$$|1,0\rangle = \frac{1}{\sqrt{2}}(|{-}{+}\rangle + |{+}{-}\rangle) \rightarrow a_{1,1,-\frac{1}{2},\frac{1}{2}} = a_{1,1,\frac{1}{2},-\frac{1}{2}} = \frac{1}{\sqrt{2}}$$

$$|1,-1\rangle = |{-}{-}\rangle \rightarrow a_{1,1,-\frac{1}{2},-\frac{1}{2}} = 1$$

are called a *triplet*.

The state $|s=0,m=0\rangle = |0,0\rangle$ can then be found as follows: the state has $m=0$, $\hat{S}_z |0,0\rangle = 0$ and thus each state(term) in the linear combination must have $m = m_1 + m_2 = 0$, which means we must be able to write

$$|0,0\rangle = a|{+}{-}\rangle + b|{-}{+}\rangle$$

where

$$|a|^2 + |b|^2 = 1 \text{ (state is normalized to 1)}$$

Now we must also have

$$\langle 1,0 \mid 0,0\rangle = 0 \text{ (since the } |s,m\rangle \text{ states are orthogonal)}$$

which implies that

$$\frac{1}{\sqrt{2}}a + \frac{1}{\sqrt{2}}b = 0 \rightarrow b = -a$$

We then have

$$2|a|^2 = 1 \rightarrow a = \frac{1}{\sqrt{2}} = -b$$

and

$$|0,0\rangle = |s=0, m=0\rangle = \frac{1}{\sqrt{2}}|+-\rangle - \frac{1}{\sqrt{2}}|-+\rangle$$

This is called a *singlet state*.

That completes the construction of the angular momentum states for the combined system of two spin-1/2 systems.

We now *generalize* this procedure for the addition of any two angular momenta.

General Addition of Two Angular Momenta

Given two angular momenta $\vec{J}_{1,op}$ and $\vec{J}_{2,op}$, we have the operators and states for each separate system

$$\vec{J}_{1,op} \rightarrow \vec{J}^2_{1,op}, \hat{J}_{1z}, \hat{J}_{1\pm} \rightarrow |j_1, m_1\rangle$$
$$\vec{J}_{2,op} \rightarrow \vec{J}^2_{2,op}, \hat{J}_{2z}, \hat{J}_{2\pm} \rightarrow |j_2, m_2\rangle$$

with

$$\vec{J}^2_{1,op}|j_1,m_1\rangle = \hbar^2 j_1(j_1+1)|j_1,m_1\rangle, \hat{J}_{1z}|j_1,m_1\rangle = \hbar m_1|j_1,m_1\rangle$$
$$\hat{J}_{1\pm}|j_1,m_1\rangle = \hbar\sqrt{j_1(j_1+1) - m_1(m_1\pm 1)}\,|j_1 \pm 1, m_1\rangle$$
$$\vec{J}^2_{2,op}|j_2,m_2\rangle = \hbar^2 j_2(j_2+1)|j_2,m_2\rangle, \hat{J}_{2z}|j_2,m_2\rangle = \hbar m_2|j_2,m_2\rangle$$
$$\hat{J}_{2\pm}|j_2,m_2\rangle = \hbar\sqrt{j_2(j_2+1) - m_2(m_2\pm 1)}\,|j_2 \pm 1, m_2\rangle$$

Remember that there are $2j_1 + 1$ possible m_1 values and $2j_2 + 1$ possible m_2 values.

Since all of the "1" operators commute with all of the "2" operators, we can find a common eigenbasis for the four operators $\vec{J}_{1,op}^2, \hat{J}_{1z}, \vec{J}_{2,op}^2, \hat{J}_{2z}$ in terms of the direct-product states

$$|j_1, j_2, m_1, m_2\rangle = |j_1, m_1\rangle \otimes |j_2, m_2\rangle$$

For the combined system we define total operators as before

$$\vec{J}_{op} = \vec{J}_{1,op} + \vec{J}_{2,op} = \text{ total angular momentum}$$
$$\hat{J}_z = \hat{J}_{1z} + \hat{J}_{2z}, \left[\vec{J}_{op}^2, \hat{J}_z\right] = 0$$
$$\left[\vec{J}_{op}^2, \hat{J}_{1,op}^2\right] = 0 \quad , \quad \left[\vec{J}_{op}^2, \hat{J}_{2,op}^2\right] = 0$$
$$\left[\hat{J}_{1,op}^2, \hat{J}_z\right] = 0 \quad , \quad \left[\hat{J}_{2,op}^2, \hat{J}_z\right] = 0$$

These commutators imply that we can construct a common eigenbasis of $\vec{J}_{1,op}^2, \hat{J}_{1z}, \vec{J}_{2,op}^2, \hat{J}_{2z}$ using the states $|j_1, j_2, m_1, m_2\rangle$ where

$$\vec{J}_{op}^2 |j, m\rangle = \hbar^2 j(j+1) |j, m\rangle \text{ and } \hat{J}_z |j, m\rangle = \hbar m |j, m\rangle$$

There are $2j + 1$ possible m values for each allowed j value.

We cannot use the operators $\hat{J}_{1z}, \hat{J}_{2z}$ to construct the eigenbasis for the combined system because they do not commute with $\vec{J}_{op}^2$.

Remember that in order for a label to appear in a ket vector it must be one of the eigenvalues of a set of commuting observables since only such a set shares a common eigenbasis.

We now determine how to write the $|j_1, j_2, m_1, m_2\rangle$ in terms of the $|j_1, m_1\rangle \otimes |j_2, m_2\rangle$ basis.

We have

$$|j_1, j_2, j, m\rangle$$
$$= \sum_{j_1'} \sum_{j_2'} \sum_{m_1} \sum_{m_2} |j_1', j_2', m_1, m_2\rangle \langle j_1', j_2', m_1, m_2 \mid j_1, j_2, j, m\rangle$$

where

$$\langle j_1', j_2', m_1, m_2 \mid j_1, j_2, j, m\rangle = \text{ Clebsch - Gordon coefficients}$$

This corresponds to inserting an identity operator of the form

$$\hat{I} = \sum_{j_1'} \sum_{j_2'} \sum_{m_1} \sum_{m_2} |j_1', j_2', m_1, m_2\rangle \langle j_1', j_2', m_1, m_2|$$

Since

$$\langle j_1', j_2', m_1, m_2| \vec{J}_{1,op}^{\,2} |j_1, j_2, j, m\rangle$$
$$= \hbar^2 j_1'(j_1' + 1) \langle j_1', j_2', m_1, m_2 \mid j_1, j_2, j, m\rangle$$
$$= \hbar^2 j_1(j_1 + 1) \langle j_1', j_2', m_1, m_2 \mid j_1, j_2, j, m\rangle$$

the Clebsch-Gordon(CG) coefficients must vanish unless $j'_1 = j_1$ and similarly unless $j'_2 = j_2$.

Thus we have

$$|j_1, j_2, j, m\rangle = \sum_{m_1} \sum_{m_2} |j_1, j_2, m_1, m_2\rangle \langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m\rangle$$

Also, as we saw earlier, since $\hat{J}_z = \hat{J}_{1z} + \hat{J}_{2z}$ we must have

$$\langle j_1, j_2, m_1, m_2| \hat{J}_z |j_1, j_2, j, m\rangle = \hbar(m_1 + m_2) \langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m\rangle$$
$$= \hbar m \langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m\rangle$$

which implies that the CG coefficients must vanish unless $m_1 + m_2 = m$.

Thus the only non-vanishing coefficients are

$$\langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m = m_1 + m_2 \rangle$$

and we can write

$$|j_1, j_2, j, m\rangle = \sum_{m_1} \sum_{m_2 = m - m_1} |j_1, j_2, m_1, m_2\rangle \langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m\rangle$$

$$= \sum_{m_1} |j_1, j_2, m_1, m_2 = m - m_1\rangle \langle j_1, j_2, m_1, m_2 = m - m_1 \mid j_1, j_2, j, m\rangle$$

For fixed j_1 and j_2 there are $2j_1 + 1$ possible m_1 values and $2j_2 + 1$ possible m_2 values.

Thus, there are $(2j_1 + 1)(2j_2 + 1)$ linearly independent states of the form

$$|j_1, j_2, m_1, m_2\rangle = |j_1, m_1\rangle \otimes |j_2, m_2\rangle$$

and hence the vector space describing the combined system is $(2j_1 + 1)(2j_2 + 1)$ - dimensional.

This says that there must be $(2j_1 + 1)(2j_2 + 1)$ states of the form $|j_1, j_2, j, m\rangle$ also.

We notice that there is only one state with $m = m_1 + m_2 = j_1 + j_2$, namely,

$$m_1 = j_1 \text{ and } m_2 = j_2$$

This state has the *maximum possible m value*.

There are two states with $m = m_1 + m_2 = j_1 + j_2 - 1$, namely,

$$m_1 = j_1 - 1, m_2 = j_2 \text{ and } m_1 = j_1, m_2 = j_2 - 1$$

and so on.

For **example**

$$j_1 = 2\ ,\ j_2 = 1 \rightarrow (2(2)+1))(2(1)+1)) = 15 \text{ states}$$

If we label these states by the m-values only (since j_1 and j_2 do not change) or $|m_1,m_2\rangle$, then we have (in this example)

$$m = 3 \rightarrow 1 \text{ state} \rightarrow |2,1\rangle$$
$$m = 2 \rightarrow 2 \text{ states} \rightarrow |2,0\rangle\ ,\ |1,1\rangle$$
$$m = 1 \rightarrow 3 \text{ states} \rightarrow |1,0\rangle\ ,\ |0,1\rangle\ ,\ |2,-1\rangle$$
$$m = 0 \rightarrow 3 \text{ states} \rightarrow |0,0\rangle\ ,\ |1,-1\rangle\ ,\ |-1,1\rangle$$
$$m = -1 \rightarrow 3 \text{ states} \rightarrow |-1,0\rangle\ ,\ |0,-1\rangle\ ,\ |-2,1\rangle$$
$$m = -2 \rightarrow 2 \text{ states} \rightarrow |-2,0\rangle\ ,\ |-1,-1\rangle$$
$$m = -3 \rightarrow 1 \text{ state} \rightarrow |-2,-1\rangle$$

for a total of 15 states.

The combined system, as we shall see by construction, has these states

$$j = 3 \rightarrow m = 3,2,1,0,-1,-2,-3 \rightarrow 7 \text{ states}$$
$$j = 2 \rightarrow m = 2,1,0,-1,-2 \rightarrow 5 \text{ states}$$
$$j = 1 \rightarrow m = 1,0,-1 \rightarrow 3 \text{ states}$$

for a total of 15 states.

The general rules, which follows from group theory, are

1. The combined system has allowed j values given by

$$j_1 + j_2 \geq j \geq |j_1 - j_2| \text{ in integer steps}$$

2. The total number of states is given by the total number of m-values for all the allowed j-values or

$$\sum_{j=|j_1-j_2|}^{j_1+j_2}(2j+1) = (2j_1+1)(2j_2+1)$$

We write the addition of two angular momenta symbolically as

$$j_1 \otimes j_2 = |j_1 - j_2| \oplus |j_1 - j_2| + 1 \oplus |j_1 - j_2| + 2 \oplus \ldots\ldots \oplus j_1 + j_2 - 1 \oplus j_1 + j_2$$

Examples

Our original special case of adding two spin = 1/2 systems gives

$$j_1 = j_2 = \frac{1}{2} \rightarrow j = 0, 1 \rightarrow \frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1 \rightarrow 4 \text{ states}$$

which is the result we found earlier.

Other cases are:

$$j_1 = j_2 = 1 \rightarrow j = 0, 1, 2 \rightarrow 1 \otimes 1 = 0 \oplus 1 \oplus 2 \rightarrow 9 \text{ states}$$

$$j_1 = 2, j_2 = 1 \rightarrow j = 1, 2, 3 \rightarrow 2 \otimes 1 = 1 \oplus 2 \oplus 3 \rightarrow 15 \text{ states}$$

$$j_1 = 2, j_2 = 3 \rightarrow j = 1, 2, 3, 4, 5 \rightarrow 2 \otimes 3 = 1 \oplus 2 \oplus 3 \oplus 4 \oplus 5 \rightarrow 35 \text{ states}$$

Actual Construction of States

Notation

1. states labeled $|7, 6\rangle$ are $|j, m\rangle$ states
2. states labeled $|3, 2\rangle_{\otimes}$ are $|m_1, m_2\rangle$ states
3. we suppress the j_1, j_2 labels everywhere

Procedure

1. choose j_1 and j_2
2. write down the direct-product $|m_1, m_2\rangle$ basis
3. determine the allowed j values
4. write down the maximum m state ($m = j_1 + j_2$); it is unique
5. the maximum m-value corresponds to the maximum j-value
6. use the lowering operator to generate all other m-states for this J-value; there are $2j + 1$, i.e.,

$$\hat{J}_- |j, m\rangle = \hbar\sqrt{j(j+1) - m(m-1)}\,|j, m-1\rangle$$

7. find the maximum m-state for the next lowest j-value; it is constructed from the same basis states as in the corresponding m-states for higher j-values; use orthonormality properties to figure out coefficients
8. repeat (6) and (7) until all j-values have been dealt with

More Detailed Examples (we must learn this process by doing it)

#1 - Individual system values

$$j_1 = j_2 = \frac{1}{2} \rightarrow m_1, m_2 = \frac{1}{2}, -\frac{1}{2} \text{ (we already did this example)}$$

The basis states are (we use the notation $|{+}{-}\rangle$ here instead of $|1/2,{-}1/2\rangle_{\otimes}$)

$$\begin{array}{lcccccccc} & |{+}{+}\rangle & , & |{+}{-}\rangle & , & |{-}{+}\rangle & , & |{-}{-}\rangle \\ m = m_1 + m_2 = & 1 & & 0 & & -1 & & 0 \end{array}$$

Construction Algebra

Allowed j-values are $j = 1, 0$

$j = 1$ has $2j + 1 = 3$ m-values $= 1, 0, -1$

$j = 0$ has $2j + 1 = 1$ m value $= 0$

$|1,1\rangle = |{+}{+}\rangle$ maximum or topmost (j, m) state is always unique

$$\hat{J}_- |1,1\rangle = \sqrt{2}\hbar\,|1,0\rangle = (\hat{J}_{1-} + \hat{J}_{2-})\,|{+}{+}\rangle = \hbar\,|{+}{-}\rangle + \hbar\,|{-}{+}\rangle$$

$$|1,0\rangle = \tfrac{1}{\sqrt{2}}\,|{+}{-}\rangle + \tfrac{1}{\sqrt{2}}\,|{-}{+}\rangle$$

$$\hat{J}_- |1,0\rangle = \sqrt{2}\hbar\,|1,-1\rangle = (\hat{J}_{1-} + \hat{J}_{2-})\left(\tfrac{1}{\sqrt{2}}\,|{+}{-}\rangle + \tfrac{1}{\sqrt{2}}\,|{-}{+}\rangle\right) = \sqrt{2}\hbar\,|{-}{-}\rangle$$

$|1,-1\rangle = |{-}{-}\rangle$

We must now have $|0,0\rangle = a\,|{+}{-}\rangle + b\,|{-}{+}\rangle$ (Rule 7) with

$$|a|^2 + |b|^2 = 1 \text{ and } \langle 1,0\,|\,,0,0\rangle = \tfrac{1}{\sqrt{2}}a + \tfrac{1}{\sqrt{2}}b = 0$$

which gives

$$a = -b = \tfrac{1}{\sqrt{2}}$$

or

$$|0,0\rangle = \tfrac{1}{\sqrt{2}}\,|{+}{-}\rangle - \tfrac{1}{\sqrt{2}}\,|{-}{+}\rangle$$

All the $j-$values are now done.

We end up with the Clebsch-Gordon coefficients.

#2 - Individual system values

$$j_1 = j_2 = 1 \rightarrow m_1, m_2 = 1, 0, -1$$

The basis states are

$$\begin{array}{lccccc} & |1,1\rangle_\otimes & |1,0\rangle_\otimes & |1,-1\rangle_\otimes & |0,1\rangle_\otimes & |0,0\rangle_\otimes \\ m = m_1 + m_2 = & 2 & 1 & 0 & 1 & 0 \end{array}$$

$$\begin{array}{lcccc} & |0,-1\rangle_\otimes & |-1,1\rangle_\otimes & |-1,0\rangle_\otimes & |-1,-1\rangle_\otimes \\ m = m_1 + m_2 = & -1 & 0 & -1 & -2 \end{array}$$

Construction Algebra

Allowed j-values are $j = 2, 1, 0$

$j = 2$ has $2j + 1 = 3$ m-values $= 2, 1, 0, -1, -2$

$j = 1$ has $2j + 1 = 3$ m-values $= 1, 0, -1$

$j = 0$ has $2j + 1 = 1$ m value $= 0$

$|2,2\rangle = |1,1\rangle_\otimes$ maximum or topmost (j, m) state is always unique

$$\hat{J}_- |2,2\rangle = 2\hbar |2,1\rangle = (\hat{J}_{1-} + \hat{J}_{2-}) |1,1\rangle_\otimes = \sqrt{2}\hbar |1,0\rangle_\otimes + \sqrt{2}\hbar |0,1\rangle_\otimes$$

$$|2,1\rangle = \tfrac{1}{\sqrt{2}} |1,0\rangle_\otimes + \tfrac{1}{\sqrt{2}} |0,1\rangle_\otimes$$

$$\begin{aligned} \hat{J}_- |2,1\rangle = \sqrt{6}\hbar |2,0\rangle &= (\hat{J}_{1-} + \hat{J}_{2-}) \left(\tfrac{1}{\sqrt{2}} |1,0\rangle_\otimes + \tfrac{1}{\sqrt{2}} |0,1\rangle_\otimes \right) \\ &= \hbar |1,-1\rangle_\otimes + 2\hbar |0,0\rangle_\otimes + \hbar |-1,1\rangle_\otimes \end{aligned}$$

$$|2,0\rangle = \tfrac{1}{\sqrt{6}} |1,-1\rangle_\otimes + \tfrac{2}{\sqrt{6}} |0,0\rangle_\otimes + \tfrac{1}{\sqrt{6}} |-1,1\rangle_\otimes$$

Continuing we have

$$|2,-1\rangle = \frac{1}{\sqrt{2}}|-1,0\rangle_{\otimes} + \frac{1}{\sqrt{2}}|0,-1\rangle_{\otimes}$$
$$|2,-2\rangle = |-1,-1\rangle_{\otimes}$$

which completes the five $j = 2$ states.

We must now have $|1,1\rangle = a|1,0\rangle_{\otimes} + b|0,1\rangle_{\otimes}$ with

$$|a|^2 + |b|^2 = 1 \text{ and } \langle 2,1|,1,1\rangle = \tfrac{1}{\sqrt{2}}a + \tfrac{1}{\sqrt{2}}b = 0$$

which gives

$$a = -b = \tfrac{1}{\sqrt{2}}$$

or

$$|1,1\rangle = \tfrac{1}{\sqrt{2}}|1,0\rangle_{\otimes} - \tfrac{1}{\sqrt{2}}|0,1\rangle_{\otimes}$$

We now find all of the $j = 1$ states

$$\hat{J}_-|1,1\rangle = \sqrt{2}\hbar|1,0\rangle = (\hat{J}_{1-} + \hat{J}_{2-})(\frac{1}{\sqrt{2}}|1,0\rangle_{\otimes} - \frac{1}{\sqrt{2}}|0,1\rangle_{\otimes})$$
$$= \hbar|1,-1\rangle_{\otimes} - \hbar|-1,1\rangle_{\otimes}$$
$$|1,0\rangle = \frac{1}{\sqrt{2}}|1,-1\rangle_{\otimes} - \frac{1}{\sqrt{2}}|-1,1\rangle_{\otimes}$$

and continuing

$$|1,-1\rangle = \frac{1}{\sqrt{2}}|0,-1\rangle_{\otimes} - \frac{1}{\sqrt{2}}|-1,0\rangle_{\otimes}$$

which completes the three $j = 1$ states.

We must now have $|0,0\rangle = a\,|1,-1\rangle_\otimes + b\,|0,0\rangle_\otimes + c\,|-1,1\rangle_\otimes$ with

$$|a|^2 + |b|^2 = 1 \text{ and } \langle 2,0\,|\,,0,0\rangle = \tfrac{1}{\sqrt{6}}a + \tfrac{2}{\sqrt{6}}b + \tfrac{1}{\sqrt{6}}c = 0$$

$$\text{and } \langle 1,0\,|\,,0,0\rangle = \tfrac{1}{\sqrt{6}}a - \tfrac{1}{\sqrt{6}}c = 0$$

which gives

$$a = -b = c = \tfrac{1}{\sqrt{3}}$$

or

$$|0,0\rangle = \tfrac{1}{\sqrt{3}}\,|1,-1\rangle_\otimes - \tfrac{1}{\sqrt{3}}\,|0,0\rangle_\otimes + \tfrac{1}{\sqrt{3}}\,|-1,1\rangle_\otimes$$

All the *j*-values are now done.

Two- and Three-Dimensional Systems

We now turn our attention to 2– and 3–dimensional systems that can be solved analytically.

In the position representation, the wave function

$$\psi(\vec{r}) = \langle \vec{r} \mid \psi \rangle$$

contains all the information that is knowable about the state of a physical system.

The equation that determines the wave function is the Schrodinger equation, which we derived from the energy eigenvalue equation

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

The general form of the Schrodinger equation in three dimensions is

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\vec{r}) + V(\vec{r})\psi(\vec{r}) = E\psi(\vec{r})$$

We will use a succession of concrete examples to elucidate the solution techniques and the physical principles that are involved.

2- and 3-Dimensional Infinite Wells

2-Dimensional Infinite Square Well - Cartesian Coordinates

The potential energy function is

$$V(x,y) = \begin{cases} 0 & |x| < \frac{a}{2} \text{ and } |y| < \frac{a}{2} \rightarrow \text{ region I} \\ \infty & \text{otherwise} \rightarrow \text{ region II} \end{cases}$$

This is a simple extension of the 1–dimensional infinite well problem, but it is useful because it illustrates all the ideas we will need for more complicated problems.

In region I
$$\frac{\partial^2\psi(x,y)}{\partial x^2} + \frac{\partial^2\psi(x,y)}{\partial y^2} = -\frac{2mE}{\hbar^2}\psi(x,y)$$

In region II
$$\psi(x,y) = 0$$

since the potential is infinite over an extended region.

We solve this equation by the separation of variables (SOV) technique.

We assume
$$\psi(x,y) = X(x)Y(y)$$

Upon substitution we get

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} = -\frac{2mE}{\hbar^2}$$

Each term on the left-hand side of the equation is a function only of a single variable and hence the only way to satisfy the equation for all x and y is to set each function of a single variable equal to a constant.

We have

$$\frac{1}{X}\frac{d^2X}{dx^2} = -\frac{2mE_x}{\hbar^2} = \text{constant}$$

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -\frac{2mE_y}{\hbar^2} = \text{constant}$$

$$E = E_x + E_y$$

If we choose

$$\frac{2mE_x}{\hbar^2} = k_x^2 \quad , \quad \frac{2mE_y}{\hbar^2} = k_y^2 \quad , \quad \frac{2mE}{\hbar^2} = k^2$$

we get the solutions

$$X(x) = A\sin k_x x + B\cos k_x x$$

$$Y(y) = C\sin k_y y + D\cos k_y y$$

with the boundary conditions

$$X\left(\pm\frac{a}{2}\right) = 0 = Y\left(\pm\frac{a}{2}\right)$$

For the function $X(x)$ each boundary condition implies two solution types

$$A = 0 \text{ or } \sin\frac{k_x a}{2} = 0 \text{ and } B = 0 \text{ or } \cos\frac{k_x a}{2} = 0$$

These are both summarized in the solution

$$X(x) = \begin{cases} \sin \frac{n_x \pi x}{a} & n_x = \text{even} \\ \\ \cos \frac{n_x \pi x}{a} & n_x = \text{odd} \end{cases} \qquad k_x = \frac{n_x \pi}{a} \qquad n_x = \text{integer}$$

and similarly for $Y(y)$

$$Y(y) = \begin{cases} \sin \frac{n_y \pi y}{a} & n_y = \text{even} \\ \\ \cos \frac{n_y \pi y}{a} & n_y = \text{odd} \end{cases} \qquad k_y = \frac{n_y \pi}{a} \qquad n_y = \text{integer}$$

The corresponding energy eigenvalues are

$$E = (n_x^2 + n_y^2)\frac{\pi^2 \hbar^2}{2ma^2} \quad , \quad n_x, n_y = 1, 2, 3, 4 \ldots$$

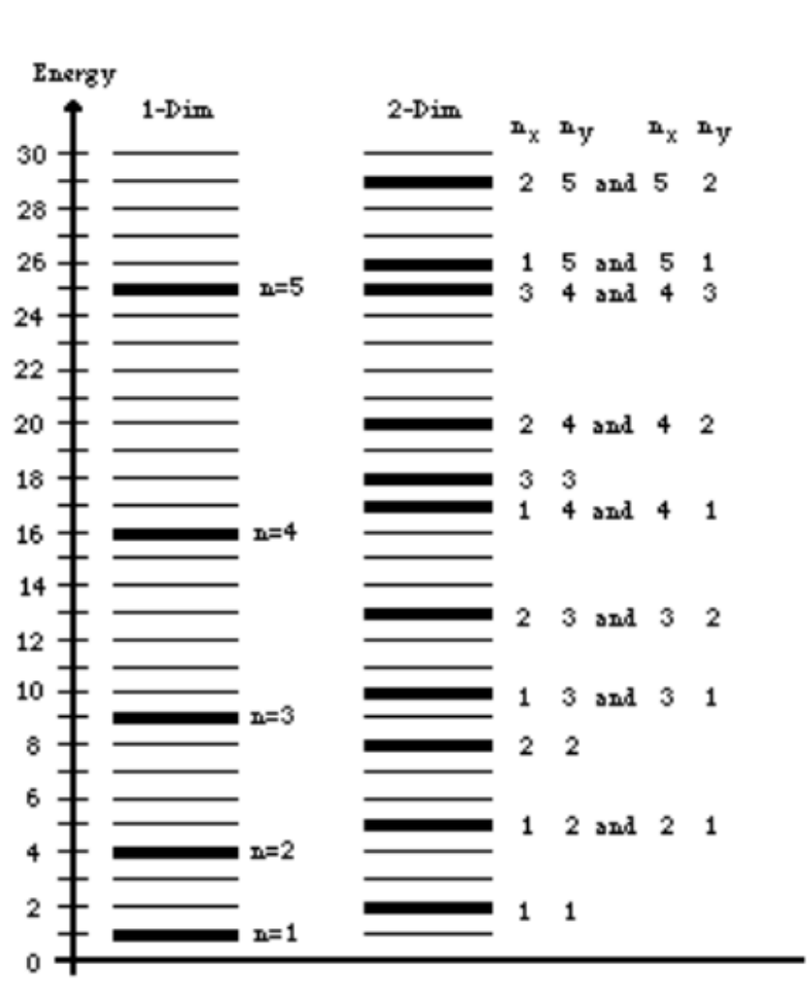
The 1–dimensional result we found earlier was

$$E_{1\,\mathrm{dim}} = \frac{n^2 \pi^2 \hbar^2}{2ma^2} \quad , n = 1, 2, 3, 4 \ldots$$

A plot of these levels for comparison is shown in the Figure;

we choose $\frac{\pi^2 \hbar^2}{2ma^2} = 1$

Comparison of 1D and 2D Infinite Wells



The major change is not only that the level structure gets more complex, but also that a major new feature appears, namely, *degeneracy*.

Several of the 2–dimensional well levels are degenerate, which means that *different sets* of quantum numbers give the *same energy eigenvalue*.

In the energy level diagram above the $E = 5,10,13,17,20,25,26$ and 29 levels are all two-fold degenerate.

This degeneracy arises from the fact that $V(x, y) = V(-x, -y)$ and hence parity is conserved.

This means that the correct physical eigenstates should be simultaneous eigenstates of both the parity and the energy.

The first three wave functions are

$$\psi_{11}(x,y) = \cos\frac{\pi x}{a}\cos\frac{\pi y}{a} \quad , \quad E = E_{11}$$

$$\psi_{12}(x,y) = \cos\frac{\pi x}{a}\sin\frac{2\pi y}{a} \quad , \quad E = E_{12}$$

$$\psi_{21}(x,y) = \sin\frac{2\pi x}{a}\cos\frac{\pi y}{a} \quad , \quad E = E_{21} = E_{12}$$

A simple calculation shows that we have

$$\langle 11 \mid 12\rangle = \langle 11 \mid 21\rangle = 0 \rightarrow \text{ orthogonal and } \langle 21 \mid 12\rangle \neq 0 \rightarrow \text{ not orthogonal}$$

We can construct two new eigenfunctions from the degenerate pair that are orthogonal using the Gram-Schmidt process.

We get

$$\psi_{12}^{+} = \psi_{12} + \psi_{21} \text{ and } \psi_{12}^{-} = \psi_{12} - \psi_{21}$$

We then have

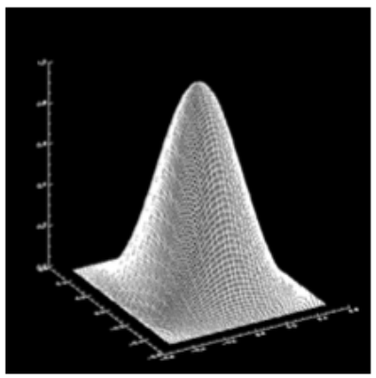
$$\langle 12+ \mid 12- \rangle = 0 \rightarrow \text{ orthogonal}$$

We also note that for the parity operator $\wp$ we have

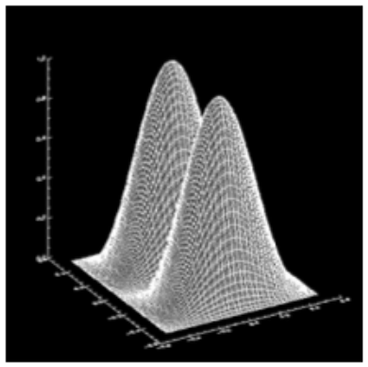
$$\hat{\wp}\,|11\rangle = |11\rangle \quad , \quad \hat{\wp}\,|12+\rangle = |12+\rangle \quad , \quad \hat{\wp}\,|12-\rangle = -|12-\rangle$$

so that the new eigenfunctions are simultaneous eigenstates of parity and energy.

Two of the wave functions are plotted below in the Figures as $|\psi|^2$.



$|\psi_{11}|^2$



$|\psi_{12}^{+}|^2$

Now let us turn to the infinite circular well in two dimensions.

Two-Dimensional Infinite Circular Well

We consider the potential in two dimensions

$$V(r) = \begin{cases} 0 & r < a \\ \infty & r > a \end{cases}$$

The Schrodinger equation in plane-polar coordinates is

$$-\frac{\hbar^2}{2m}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}\right]\psi(r,\varphi)=E\psi(r,\varphi)\qquad r<a$$

$$\psi(r,\varphi)=0\qquad r<a$$

We assume(SOV)

$$\psi(r,\varphi)=R(r)\Phi(\varphi)$$

which gives

$$-\frac{\hbar^2}{2m}\left[\frac{1}{rR}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right)+\frac{1}{r^2}\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2}\right]=E\qquad r<a$$

We choose a separation constant

$$\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2}=-\alpha^2\rightarrow\Phi(\varphi)=B\sin(\alpha\varphi+\delta)$$

The requirement of single-valuedness under a ϕ-rotation of 2π says that

$$\sin(\alpha\varphi+\delta)=\sin(\alpha\varphi+\delta+\alpha\pi)$$

$$\rightarrow\alpha=\text{ integer }=0,1,2,3\ldots$$

Alternatively, we could write

$$\Phi(\varphi)=Be^{i\alpha\varphi}$$

$$\alpha=\text{ integer }=\ldots-3,-2,-1,0,1,2,3,\ldots$$

Substitution of this solution leaves the radial differential equation

$$r^2\frac{d^2R}{dr^2}+r\frac{dR}{dr}+\left[\lambda^2r^2-\alpha^2\right]R=0\text{ where }\lambda^2=\frac{2mE}{\hbar^2}$$

This is Bessel's equation.

The general solution is

$$R(r) = NJ_\alpha(\lambda r) + MY_\alpha(\lambda r)$$

Now $Y_\alpha(\lambda r) \to \infty$ as $r \to 0$.

Therefore, in order to have a normalizable solution in the region $r < a$ (which includes $r = 0$), we must choose $M = 0$ and thus we have

$$R(r) = NJ_\alpha(\lambda r)$$

and the complete solution is then

$$\psi_{k\alpha}(r, \varphi) = R(r)\Phi(\varphi) = NJ_\alpha(\lambda r)e^{i\alpha\varphi}$$

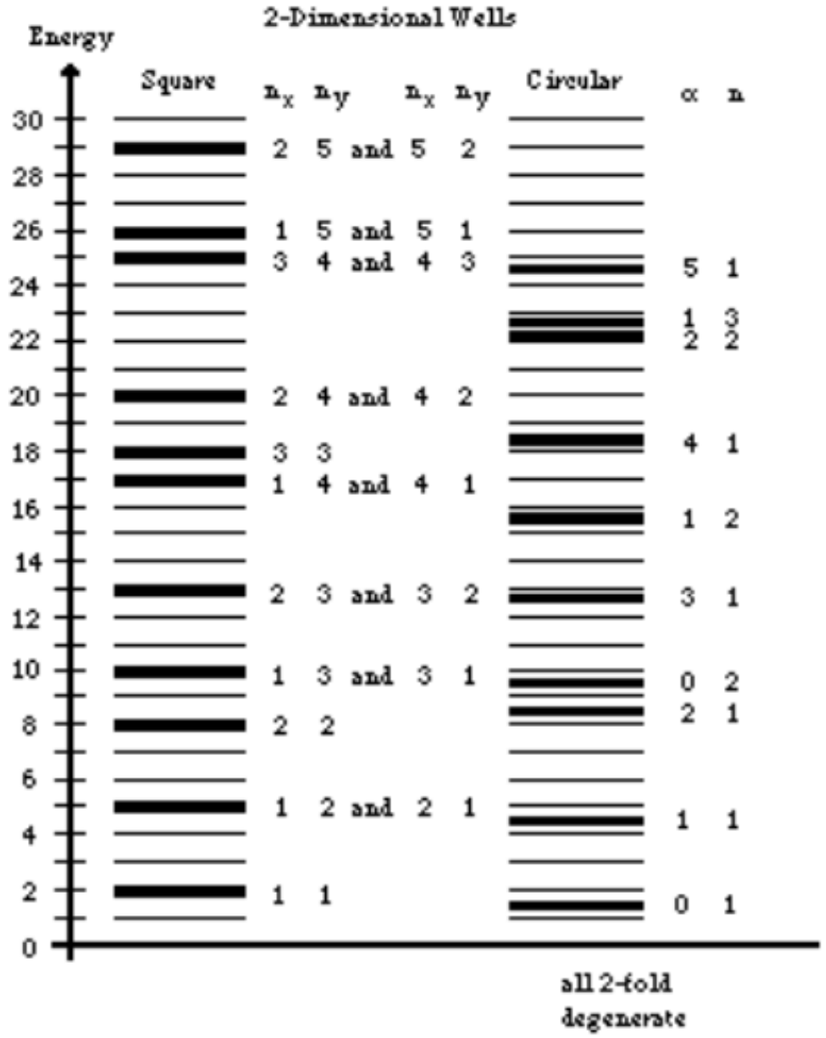
The continuity (or boundary) condition at $r = a$ is

$$\psi_{k\alpha}(a, \varphi) = 0 \to R(a) = 0 \to J_\alpha(\lambda a) = 0$$

Thus, the allowed values of λ and hence the allowed values of E are given by

$$\lambda_{n\alpha}a = z_{n\alpha} = \text{ the } n^{th} \text{ zero of } J_\alpha \to E_{n\alpha} = \frac{\hbar^2}{2ma^2}z_{n\alpha}^2$$

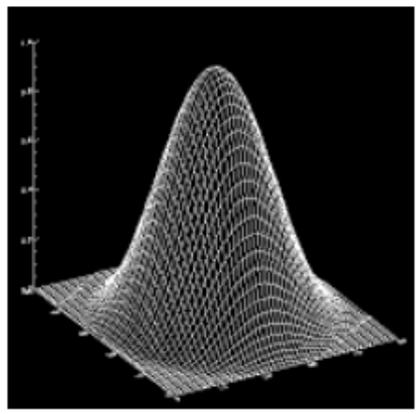
We compare the infinite square and circular wells in 2–dimensions using the energy level diagram in the Figure.



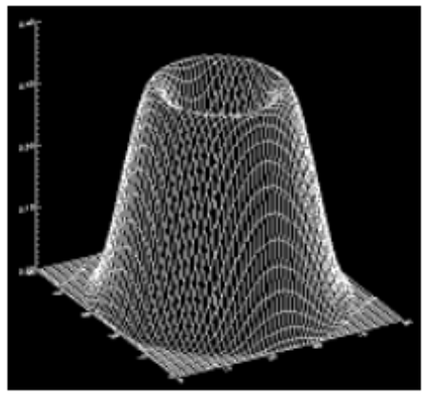
Two Dimensional Wells

Note the rather dramatic differences in both the location and degeneracies for the two sets of energy levels.

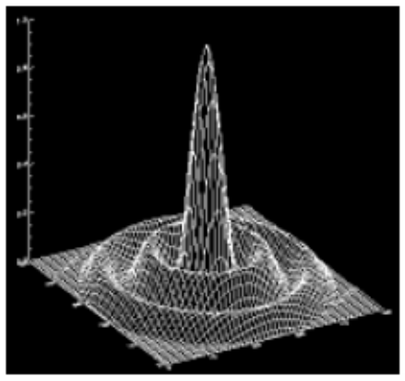
Some of the wave functions for the 2–dimensional circular well are shown in the Figures below.



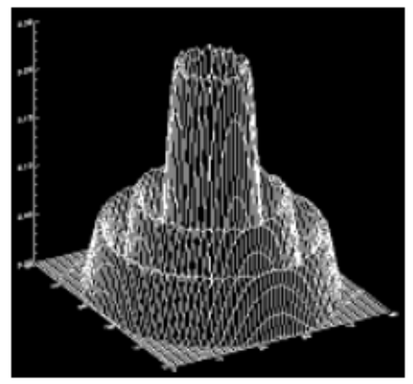
$|\psi_{01}|^2$



$|\psi_{11}|^2$



$|\psi_{03}|^2$



$|\psi_{23}|^2$

The 3–*dimensional infinite square well* is a simple extension of the 2–dimensional infinite square well.

The result for the energies is

$$E = (n_x^2 + n_y^2 + n_z^2)\frac{\pi^2\hbar^2}{2ma^2} \quad , \quad n_x, n_y, n_z = 1, 2, 3, 4\ldots$$

The 3–*dimensional infinite spherical well* involves the potential

$$V(r,\theta,\varphi) = \begin{cases} 0 & r < a \quad \text{region I} \\ \infty & r > a \quad \text{region II} \end{cases}$$

The Schrodinger equation is:

Region I

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2} = -\frac{2m}{\hbar^2}E$$

Region II

$$\psi(r,\theta,\varphi) = 0$$

The equation in region I can be rewritten in terms of the $\vec{L}^2_{op}$ operator as

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) - \frac{\vec{L}^2_{op}\psi}{\hbar^2 r^2} = -\frac{2m}{\hbar^2}E\psi$$

Since the potential energy is spherically symmetric, we have

$$\left[\vec{L}^2_{op}, \hat{H}\right] = 0$$

and thus the operators $\vec{L}^2_{op}$ and $\hat{H}$ have a common eigenbasis.

Earlier, we found the eigenfunctions of $\vec{L}^2_{op}$ to be the spherical harmonics $Y_{\ell m}(\theta,\phi)$, where

$$\vec{L}^2_{op}Y_{\ell m}(\theta,\varphi) = \hbar^2\ell(\ell+1)Y_{\ell m}(\theta,\varphi)$$

Therefore, we can write(SOV)

$$\psi(r,\theta,\varphi) = R_\ell(r)Y_{\ell m}(\theta,\varphi)$$

Substitution of this form of the solution gives the radial equation in region I

$$\frac{d^2R_\ell}{dr^2} + \frac{2}{r}\frac{dR_\ell}{dr} - \frac{\ell(\ell+1)}{r^2}R_\ell + k^2R_\ell = 0$$

where

$$E = \frac{h^2k^2}{2m}$$

In region II we have $R_\ell(r) = 0$.

The most general solution of the radial equation(which is a different form of Bessel's equation) is

$$R_\ell(r) = Aj_\ell(kr) + B\eta_\ell(kr)$$

where

$$j_\ell(kr) = \left(\frac{\pi}{2kr}\right)^{1/2} J_{\ell+1/2}(kr)$$

$$\eta_\ell(kr) = (-1)^{\ell+1}\left(\frac{\pi}{2kr}\right)^{1/2} J_{-\ell-1/2}(kr)$$

are the spherical Bessel functions.

Now $\eta_\ell(kr) \to \infty$ as $r \to 0$.

Therefore, the normalizable solution in region I (which contains $r = 0$) is

$$R_\ell(r) = Aj_\ell(kr)$$

The first few of these functions

$$j_0(x) = \frac{\sin x}{x}$$

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x}$$

$$j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x}\right)\sin x - \frac{3}{x^2}\cos x$$

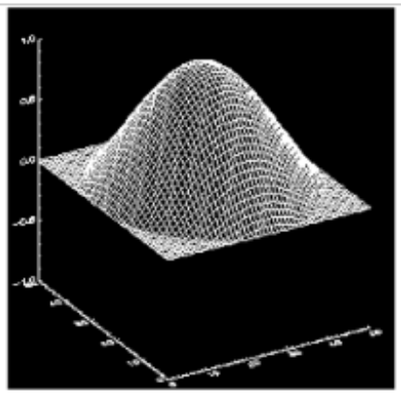
The boundary conditions give

$$j_\ell(k_{n\ell}a) = 0 \text{ where } k_{n\ell} = \text{ the } n^{th} \text{ zero of } j_\ell$$

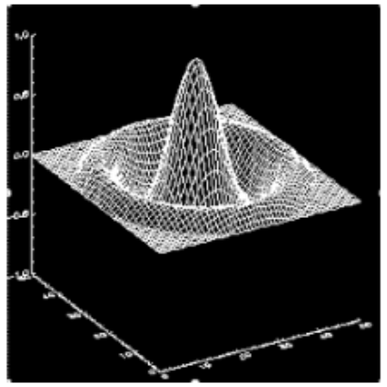
The full solution is

$$\psi_{n\ell m}(\theta,\varphi) = R_{n\ell}(r)Y_{\ell m}(\theta,\varphi) = j_\ell(k_n r)P_\ell^m(\cos\theta)e^{im\varphi}$$

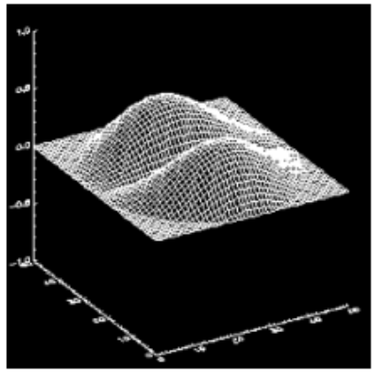
Some of the wavefunctions (absolute squares) are shown in the Figures below.



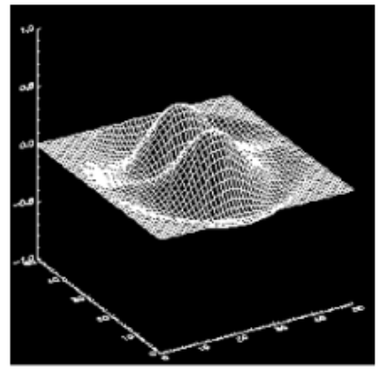
$n\ell m = 100$



$n\ell m = 200$



$n\ell m = 111$



$n\ell m = 211$

3-Dimensional Finite Well

We now consider the potential function in three dimensions

$$V(r) = \begin{cases} -V_0 & r \le a \quad \text{region I} \\ 0 & r > a \quad \text{region II} \end{cases}$$

The Schrodinger equation is

Region I (same as infinite well)

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2} + \frac{2m}{\hbar^2}V_0\psi = -\frac{2m}{\hbar^2}E\psi$$

Region II

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2}=-\frac{2m}{\hbar^2}E\psi$$

If we choose $E=-|E|<0$ for bound states, we have

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)-\frac{\vec{L}_{op}^2\psi}{\hbar^2r^2}+\frac{2m}{\hbar^2}V_0\psi=\frac{2m}{\hbar^2}|E|\psi \text{ in region I}$$

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)-\frac{\vec{L}_{op}^2\psi}{\hbar^2r^2}=\frac{2m}{\hbar^2}|E|\psi \text{ in region II}$$

We then write, as before,

$$\psi(\theta,\varphi)=R_\ell(r)Y_{\ell m}(\theta,\varphi)$$

and get

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)-\frac{\ell(\ell+1)R}{r^2}+\frac{2m}{\hbar^2}V_0R=\frac{2m}{\hbar^2}|E|R \quad \text{in region I}$$

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)-\frac{\ell(\ell+1)R}{r^2}=\frac{2m}{\hbar^2}|E|R \quad \text{in region II}$$

which finally becomes

$$\frac{d^2R}{d\rho^2}+\frac{2}{\rho}\frac{dR}{d\rho}+\left[1-\frac{\ell(\ell+1)}{\rho^2}\right]R=0$$

$$\rho=\alpha r \quad , \quad \alpha^2=\frac{2m}{\hbar^2}(V_0-|E|) \quad \text{in region I}$$

and

$$\frac{d^2R}{d\gamma^2} + \frac{2}{\gamma}\frac{dR}{d\gamma} + \left[1 - \frac{\ell(\ell+1)}{\gamma^2}\right]R = 0$$

$$\gamma = i\beta r \quad , \quad \beta^2 = \frac{2m}{\hbar^2}|E| \qquad \text{in region II}$$

The solutions are

$$R(r) = Aj_\ell(\alpha r) \qquad\qquad r \le a$$

$$R(r) = Bh_\ell^{(1)}(i\beta r) = B\left[j_\ell(i\beta r) + i\eta_\ell(i\beta r)\right] \qquad r > a$$

where

$$j_0(x) = \frac{\sin x}{x}$$

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x}$$

$$j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x}\right)\sin x - \frac{3}{x^2}\cos x$$

$$j_\ell(x) = x^\ell\left(-\frac{1}{x}\frac{d}{dx}\right)^\ell \frac{\sin x}{x}$$

which are *spherical Bessel functions of the 1st kind*, and

$$\eta_0(x) = -\frac{\cos x}{x}$$

$$\eta_1(x) = -\frac{\cos x}{x^2} - \frac{\sin x}{x}$$

$$\eta_2(x) = -\left(\frac{3}{x^3} - \frac{1}{x}\right)\cos x - \frac{3}{x^2}\sin x$$

$$\eta_\ell(x) = -x^\ell\left(-\frac{1}{x}\frac{d}{dx}\right)^\ell \frac{\cos x}{x}$$

which are *spherical Bessel functions of the 2nd kind*, and

$$h_0^{(1)}(ix) = -\frac{1}{x}e^{-x}$$

$$h_1^{(1)}(ix) = i\left(\frac{1}{x} + \frac{1}{x^2}\right)e^{-x}$$

$$h_2^{(1)}(ix) = \left(\frac{1}{x} + \frac{3}{x^2} + \frac{3}{x^3}\right)e^{-x}$$

$$h_\ell^{(1)}(ix) = j_\ell(ix) + i\eta_\ell(ix)$$

which are *spherical Hankel functions of the 1st kind*.

There is another independent solution for r > a, namely,

$$h_\ell^{(2)}(ix) = j_\ell(ix) - i\eta_\ell(ix)$$

which is a *spherical Hankel function of the 2nd kind*, but we must exclude it because it behaves like $e^{\beta r}$ as $r \to \infty$ and, hence, is not normalizable

We have also excluded $\eta_\ell(\alpha r)$ from the solution for $r \le a$ because it diverges at $r = 0$.

We note for future reference that we have the asymptotic behaviors

$$j_\ell(x) \underset{x\to 0}{\to} \frac{x^\ell}{1\cdot 3\cdot 5\cdots(2\ell+1)} \text{ and } \eta_\ell(x) \underset{x\to 0}{\to} \frac{1\cdot 3\cdot 5\cdots(2\ell-1)}{x^\ell}$$

$$j_\ell(x) \underset{x\to\infty}{\to} \frac{1}{x}\cos\left[x - \frac{(\ell+1)\pi}{2}\right] \text{ and } \eta_\ell(x) \underset{x\to\infty}{\to} \frac{1}{x}\sin\left[x - \frac{(\ell+1)\pi}{2}\right]$$

$$h_\ell^{(1)}(x) \underset{x\to\infty}{\to} \frac{1}{x}e^{i\left[x - \frac{1}{2}(\ell+1)\pi\right]} \text{ and } h_\ell^{(2)}(x) \underset{x\to\infty}{\to} \frac{1}{x}e^{-i\left[x - \frac{1}{2}(\ell+1)\pi\right]}$$

Since both R and dR/dr are continuous at $r = a$, we can combine the two continuity equations into one using the continuity of the so-called logarithmic derivative

$$\frac{1}{R}\frac{dR}{dr} \text{ at } r = a$$

For *each* value of ℓ this gives a transcendental equation for the energy E.

Examples

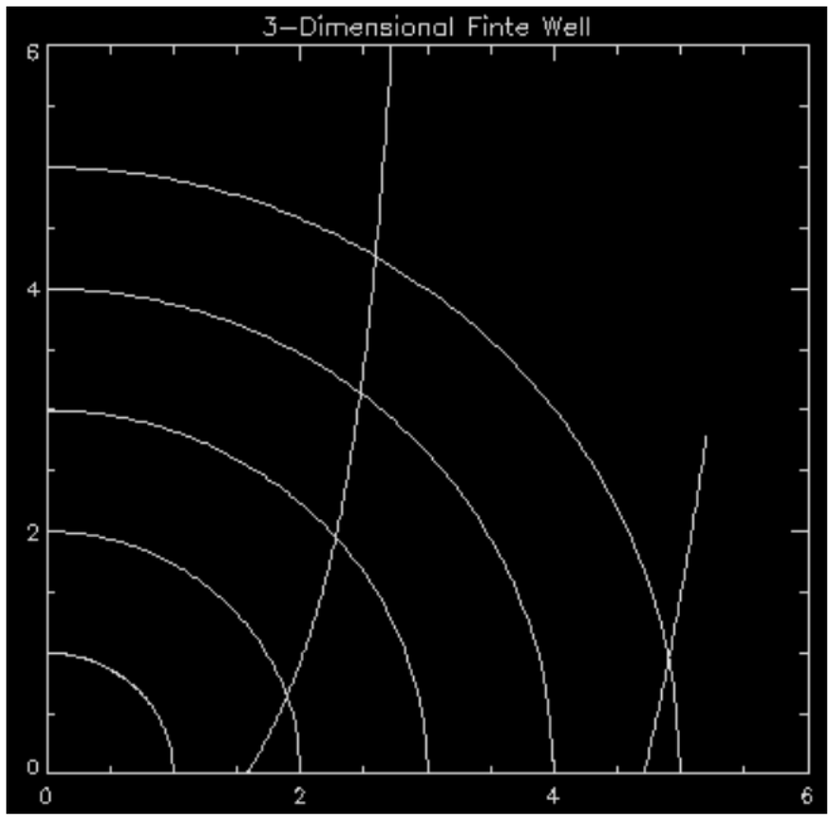
$$\ell = 0 \qquad \xi \cot \xi = -\zeta \quad \text{and} \quad \xi^2 + \zeta^2 = \frac{2mV_0a^2}{\hbar^2}$$

$$\xi = \alpha a \quad \text{and} \quad \zeta = \beta a$$

$$\ell = 1 \qquad \frac{\cot \xi}{\xi} - \frac{1}{\xi^2} = \frac{1}{\zeta} + \frac{1}{\zeta^2} v\xi^2 + \zeta^2 = \frac{2mV_0a^2}{\hbar^2}$$

$$\xi = \alpha a \quad \text{and} \quad \zeta = \beta a$$

A graphical solution for the $\ell = 0$ case is shown in the Figure.



Graphical Solution

For a given well, only one circle exists on the plot.

The solutions (energy eigenvalues) are given by the intersection of that circle with the cotangent curves.

The big change from the finite well in one dimension is that the quantity

$$\frac{2mV_0a^2}{\hbar^2}$$

must be larger than some minimum value before any bound state exists (this corresponds to the radius of the smallest circle that intersects the cotangent curves).

In particular,

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{\pi}{2}\right)^2 \rightarrow \text{ no solution}$$
$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{3\pi}{2}\right)^2 \rightarrow \text{ 1 solution}$$
$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{5\pi}{2}\right)^2 \rightarrow \text{ 2 solutions}$$

Two-Dimensional Harmonic Oscillator

The 2-dimensional harmonic oscillator system has the Hamiltonian

$$\begin{aligned}\hat{H} &= \frac{1}{2m}\left(\hat{p}_x^2 + \hat{p}_y^2\right) + \frac{1}{2}m\omega^2(\hat{x}^2 + \hat{y}^2) \\ &= \left(\frac{\hat{p}_x^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2\right) + \left(\frac{\hat{p}_y^2}{2m} + \frac{1}{2}m\omega^2\hat{y}^2\right) \\ &= \hat{H}_x + \hat{H}_y\end{aligned}$$

where we have the commutation relations

$$\left[\hat{H}_x, \hat{H}_y\right] = 0 = \left[\hat{H}, \hat{H}_x\right] = \left[\hat{H}, \hat{H}_y\right]$$

These commutators imply that $\hat{H}$, $\hat{H}_x$ and $\hat{H}_y$ have a common eigenbasis.

We label their common state vectors by

$$|E\rangle = |E_x, E_y\rangle = |E_x\rangle\,|E_y\rangle$$

with

$$\hat{H}_x\,|E_x\rangle = E_x\,|E_x\rangle$$

$$\hat{H}_y\,|E_y\rangle = E_y\,|E_y\rangle$$

$$\hat{H}\,|E\rangle = E\,|E\rangle = (\hat{H}_x + \hat{H}_y)\,|E_x\rangle\,|E_y\rangle$$

$$= (E_x + E_y)\,|E\rangle$$

or

$$E = E_x + E_y$$

Now $\hat{H}_x$ (and $\hat{H}_y$) each represent a 1-dimensional oscillator.

This suggests that we define new operators for the x coordinate

$$a_x = \sqrt{\frac{m\omega}{2\hbar}}\hat{x} + \frac{i}{\sqrt{2m\hbar\omega}}\hat{p}_x = (a_x^+)^+$$

where

$$[\hat{x}, \hat{p}_x] = i\hbar \rightarrow [\hat{a}_x, a_x^+] = 1$$

and

$$\hat{H}_x = \hbar\omega\left(\hat{a}_x^+\hat{a}_x + \frac{1}{2}\right) = \hbar\omega\left(\hat{N}_x + \frac{1}{2}\right)$$

As we found earlier, $\hat{N}_x$ has an eigenvalue equation

$$\hat{N}_x\,|n_x\rangle = n_x\,|n_x\rangle \quad , \quad n_x = 0, 1, 2, 3, \ldots\ldots$$

We then have

$$\hat{H}_x |n_x\rangle = \hbar\omega \left(\hat{N}_x + \frac{1}{2}\right)|n_x\rangle = \hbar\omega \left(n_x + \frac{1}{2}\right)|n_x\rangle$$

or

$$E_x = \hbar\omega \left(n_x + \frac{1}{2}\right)$$

In a similar manner, we repeat this process for the *y* coordinate

$$a_y = \sqrt{\frac{m\omega}{2\hbar}}\hat{y} + \frac{i}{\sqrt{2m\hbar\omega}}\hat{p}_y = \left(a_y^+\right)^+$$

$$[\hat{y}, \hat{p}_y] = i\hbar \rightarrow \left[\hat{a}_y, a_y^+\right] = 1$$

$$\hat{H}_y = \hbar\omega \left(\hat{a}_y^+ \hat{a}_y + \frac{1}{2}\right) = \hbar\omega \left(\hat{N}_y + \frac{1}{2}\right)$$

As we found earlier, $\hat{N}_y$ has an eigenvalue equation

$$\hat{N}_y |n_y\rangle = n_y |n_y\rangle \quad , \quad n_y = 0, 1, 2, 3, \ldots\ldots\ldots$$

We then have

$$\hat{H}_y |n_y\rangle = \hbar\omega \left(\hat{N}_y + \frac{1}{2}\right)|n_y\rangle = \hbar\omega \left(n_y + \frac{1}{2}\right)|n_y\rangle$$

or

$$E_y = \hbar\omega \left(n_y + \frac{1}{2}\right)$$

Putting this all together we get

$$E = E_x + E_y = \hbar\omega(n_x + n_y + 1) = \hbar\omega(n + 1)$$

The table below gives the resulting energy level structure.

n_x	n_y	$E/\hbar\omega$	n
0	0	1	0
1	0	2	1
0	1	2	1
0	2	3	2
2	0	3	2
1	1	3	2

Energy Levels - 2D Oscillator

Each energy value, which is characterized by the quantum number n, has a degeneracy equal to $(n + 1)$.

The existence of degeneracy indicates (this is a general rule) that there is another operator that commutes with $\hat{H}$.

Since this is a central force in the x–y plane, it is not difficult to guess that the other operator that commutes with $\hat{H}$ is the angular momentum about the axis perpendicular to the plane (the z–axis), $\hat{L}_z$.

Since all of the x-type operators commute with all of the y-type operators we can write

$$\hat{L}_z = (\vec{r}_{op} \times \vec{p}_{op})_z = \hat{x}\hat{p}_y - \hat{y}\hat{p}_x$$

Inverting the standard operator definitions we have

$$\hat{x} = \sqrt{\frac{\hbar}{2m\omega}}(\hat{a}_x + \hat{a}_x^+) \quad , \quad \hat{p}_x = \frac{1}{i}\sqrt{\frac{m\hbar\omega}{2}}(\hat{a}_x - \hat{a}_x^+)$$

$$\hat{y} = \sqrt{\frac{\hbar}{2m\omega}}(\hat{a}_y + \hat{a}_y^+) \quad , \quad \hat{p}_y = \frac{1}{i}\sqrt{\frac{m\hbar\omega}{2}}(\hat{a}_y - \hat{a}_y^+)$$

which gives

$$\hat{L}_z = \frac{\hbar}{i}(\hat{a}_x^+ \hat{a}_y - \hat{a}_y^+ a_x)$$

Now using

$$\left[\hat{a}_x, \hat{N}_x\right] = \hat{a}_x \quad , \quad \left[\hat{a}_x^+, \hat{N}_x\right] = -\hat{a}_x^+$$
$$\left[\hat{a}_y, \hat{N}_y\right] = \hat{a}_y \quad , \quad \left[\hat{a}_y^+, \hat{N}_y\right] = -\hat{a}_y^+$$

we get

$$\left[\hat{H}_x, \hat{L}_z\right] = \frac{\hbar}{i}(\hat{a}_y \hat{a}_x^+ + \hat{a}_x \hat{a}_y^+) = -\left[\hat{H}_y, \hat{L}_z\right]$$

or

$$\left[\hat{H}, \hat{L}_z\right] = 0$$

Therefore, $\hat{H}$ and $\hat{L}_z$ share a common eigenbasis.

This new eigenbasis will not be an eigenbasis for $\hat{H}_x$ or $\hat{H}_y$ separately since they do not commute with $\hat{L}_z$.

This suggests that we use linear combinations of the degenerate eigenstates to find the eigenstates of $\hat{L}_z$.

This works because linear combinations for fixed n remain eigenstates of $\hat{H}$ (they are no longer eigenstates of $\hat{H}_x$ or $\hat{H}_y$ however).

We define the eigenstates and eigenvalues of $\hat{L}_z$ by the equation

$$\hat{L}_z |m\rangle = m\hbar |m\rangle$$

and the common eigenstates of $\hat{H}$ and $\hat{L}_z$ by $|n,m\rangle$ where

$$\hat{H}\,|n,m\rangle = \hbar\omega(n+1)\,|n,m\rangle$$
$$\hat{L}_z\,|n,m\rangle = m\hbar\,|n,m\rangle$$

For notational clarity we will write the old states as $|n_x\rangle|n_y\rangle$.

For the $n = 0$ states (there is only one) we have

$$\hat{L}_z\,|0,0\rangle = \hat{L}_z\,|0\rangle\,|0\rangle = 0 \rightarrow m = 0\ ,\ n = 0$$

Now we look at the $n = 1$ states.

We let $$|\psi\rangle = a\,|0\rangle\,|1\rangle + b\,|1\rangle\,|0\rangle \ \text{ where } \ |a|^2 + |b|^2 = 1 \ \text{(normalization)}$$

Since the two states that make up the linear combination are both eigenstates of $\hat{H}$ with $n = 1$, the linear combination is an eigenstate of $\hat{H}$ with $n = 1$ for any choice of a and b.

We therefore choose a and b to make this state an eigenstate of $\hat{L}_z$.

We must have

$$\begin{aligned}\hat{L}_z\,|\psi\rangle &= \hat{L}_z\,|1,m\rangle = \hat{L}_z(a\,|0\rangle\,|1\rangle + b\,|1\rangle\,|0\rangle)\\ &= \frac{\hbar}{i}(\hat{a}_x^{+}\hat{a}_y - \hat{a}_y^{+}a_x)(a\,|0\rangle\,|1\rangle + b\,|1\rangle\,|0\rangle)\\ &= m\hbar(a\,|0\rangle\,|1\rangle + b\,|1\rangle\,|0\rangle)\end{aligned}$$

Using

$$\hat{a}\,|n\rangle = \sqrt{n}\,|n-1\rangle \quad , \quad \hat{a}^{+}\,|n\rangle = \sqrt{n+1}\,|n+1\rangle$$

we get

$$a\frac{\hbar}{i}\,|1\rangle\,|0\rangle - b\frac{\hbar}{i}\,|0\rangle\,|1\rangle = m\hbar(a\,|0\rangle\,|1\rangle + b\,|1\rangle\,|0\rangle)$$

or

$$ma = ib \quad and \quad mb = -ia$$

Dividing these two equations we get

$$\frac{a}{b} = -\frac{b}{a}$$

which implies that

$$a^2 = -b^2 \rightarrow a = \pm ib$$

and

$$m = -\frac{b}{ia} = \begin{cases} +1 & a = +ib \\ -1 & a = -ib \end{cases}$$

Normalization then says that

$$m = \begin{cases} +1 & a = \frac{1}{\sqrt{2}}, b = -\frac{i}{\sqrt{2}} \\ -1 & a = \frac{1}{\sqrt{2}}, b = +\frac{i}{\sqrt{2}} \end{cases}$$

or

$$|1, \pm 1\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle |1\rangle \mp |1\rangle |0\rangle \right)$$

This gives a new characterization of the first two excited energy levels as shown in the Table.

n	m	$E/\hbar\omega$
0	0	1
1	+1	2
1	-1	2

Energy Levels - (n,m) Characterization

Let us now do the $n = 2$ states.

In the same way we assume

$$|\psi\rangle = a |2\rangle |0\rangle + b |1\rangle |1\rangle + c |0\rangle |2\rangle$$

where normalization gives

$$|a|^2 + |b|^2 + |c|^2 = 1$$

We then have

$$\begin{aligned}
\hat{L}_z |\psi\rangle &= \hat{L}_z |2,m\rangle = m\hbar |\psi\rangle \\
&= m\hbar a |2\rangle |0\rangle + m\hbar b |1\rangle |1\rangle + m\hbar c |0\rangle |2\rangle \\
&= \frac{\hbar}{i}(\hat{a}_x^+ \hat{a}_y - \hat{a}_y^+ \hat{a}_x)(a |2\rangle |0\rangle + b |1\rangle |1\rangle + c |0\rangle |2\rangle) \\
&= \frac{\hbar}{i}\left[\sqrt{2}b |2\rangle |0\rangle + \sqrt{2}(c-a) |1\rangle |1\rangle - \sqrt{2}b |0\rangle |2\rangle\right]
\end{aligned}$$

which gives

$$ma = -i\sqrt{2}b$$

$$mc = +i\sqrt{2}b$$

$$mb = -i\sqrt{2}(c-a)$$

$$\frac{a}{c} = -1 \rightarrow c = -a$$

$$\frac{c}{b} = -\frac{b}{c-a} \rightarrow \frac{a}{b} = -\frac{b}{2a} \rightarrow b^2$$

$$= -2a^2 \rightarrow b = \pm i\sqrt{2a}$$

Putting these pieces all together we have

$$a = \frac{1}{2} = -c \quad , \quad b_\pm = \pm i\frac{\sqrt{2}}{2}$$

This implies the *m*-values

$$m = \begin{cases} \frac{\sqrt{2}b_+}{ia} & +2 \\ \frac{\sqrt{2}b_-}{ia} & -2 \end{cases}$$

or

$$|2,0\rangle = \frac{1}{\sqrt{2}}|2\rangle|0\rangle + \frac{1}{\sqrt{2}}|0\rangle|2\rangle$$

Thus, the final energy levels are as shown in the Table.

n	m	$E/\hbar\omega$
0	0	1
1	+1	2
1	-1	2
2	+2	3
2	0	3
2	-2	3

Energy Levels - (n,m) Characterization

What a strange result?

The allowed *m*-values are separated by $\Delta m = \pm 2$.

Let us look at this system using the Schrodinger equation to help us understand what is happening.

We have (using plane-polar coordinates)

$$-\frac{h^2}{2M}\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2\psi}{\partial\varphi^2}\right)+\frac{1}{2}M\omega^2r^2\psi = E\psi$$

Choosing

$$\psi(r,\varphi) = R(r)\Phi(\varphi)$$

we get

$$-\frac{h^2}{2M}\frac{1}{R}\left(\frac{d^2R}{dr^2}+\frac{1}{r}\frac{dR}{dr}\right)-\frac{h^2}{2Mr^2}\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2}+\frac{1}{2}M\omega^2r^2=E$$

Now we must have

$$\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2}=-m^2=\text{constant}$$

which produces a radial equation of the form

$$-\frac{h^2}{2M}\left(\frac{d^2R}{dr^2}+\frac{1}{r}\frac{dR}{dr}\right)+\left(\frac{h^2m^2}{2Mr^2}+\frac{1}{2}M\omega^2r^2\right)R=ER$$

Now we change the variables using

$$r=\rho y \quad , \quad \rho=\frac{\hbar}{M\omega}$$

and get

$$\frac{d^2R}{dy^2}+\frac{1}{y}\frac{dR}{dy}+\left(\varepsilon-y^2-\frac{m^2}{y^2}\right)R=0$$

where

$$\varepsilon=\frac{E}{\hbar\omega}$$

As with the solution of other differential equations, the procedure to follow is to extract out the asymptotic behavior as $y \to 0, \infty$ and solve the equation for the remaining function by recognizing the well-known equation that results.

As $y \to \infty$ the dominant term will be y^2 and the equation for this behavior is

$$\frac{d^2R}{dy^2}-y^2R=0$$

which has a solution $R \to e^{-y^2/2}$.

As $y \to 0$ the dominant term will be $1/y^2$ and the equation for this behavior is

$$\frac{d^2R}{dy^2} - \frac{m^2}{y^2}R = 0$$

which has a solution $R \to y^{|m|}$, where we have excluded any negative powers since that solution would diverge at $y = 0$.

Therefore we assume $R = y^{|m|}e^{-y^2/2}G(y)$.

Substitution gives the equation for G as

$$\frac{d^2G}{dy^2} + \left(\frac{2|m|+1}{y} - 2y\right)\frac{dG}{dy} + (\varepsilon - 2 - 2|m|)\,G = 0$$

Changing the variable again to $z = y^2$ we have

$$\frac{d^2G}{dz^2} + \left(\frac{|m|+1}{z} - 1\right)\frac{dG}{dz} + \left(\frac{\varepsilon - 2(|m|+1)}{4z}\right)G = 0$$

If we are clever, we recognize this as Laguerre's equation.

If not, we make a series solution substitution

$$G(z) = \sum_{s=0}^{\infty} b_s z^s$$

which gives the recursion relation

$$\frac{b_{s+1}}{b_s} = \frac{s + \frac{|m|+1}{2} - \frac{\varepsilon}{4}}{(s+1)(s+|m|+1)} \to \frac{1}{s} \text{ for large } s$$

This says that unless the series terminates (becomes a polynomial in z) it will behave like $e^z = e^{y^2}$ which implies that the solution for $R(y)$ will diverge for large y and thus, not be normalizable.

If we choose the maximum s-value to be $s_{max} = n_r$, then we can terminate the series by choosing

$$\varepsilon = 2|m| + 2 + 4n_r$$

which then gives us the allowed energy eigenvalues

$$E_{n_r,m} = \hbar\omega(|m| + 2n_r + 1)$$

The polynomial solutions are the generalized Laguerre polynomials $L^{|m|}(z)$.

The first few Laguerre polynomials are

$$L_0^k(z) = 1 \quad , \quad L_1^k(z) = 1 + k - z$$

$$L_2^k(z) = \frac{1}{2}(2 + 3k + k^2 - 2z(k+2) + z^2)$$

The full wave function is

$$\psi_{n_r,m}(r,\varphi) = r^{|m|} e^{-\frac{r^2}{2\rho^2}} L_{n_r}^{|m|}\left(\frac{r^2}{\rho^2}\right) e^{im\varphi}$$

and the energy level structure is shown in the Table.

which is the same structure (with different labels) as in the operator solution.

$E/\hbar\omega$	n_r	m	degeneracy
1	0	0	1
2	0	+1	2
2	0	-1	
3	0	+2	3
3	1	0	
3	0	-2	
4	0	+3	4
4	1	+1	
4	1	-1	
4	0	-3	

Energy Levels - 2D Oscillator

The fact that $\Delta m = 2$ in the 2-dimensional case is one of many peculiarities associated with two dimensions that does not appear three dimensions.

What happens in 3 dimensions?

In Cartesian coordinates we have a simple extension of the 2-dimensional case.

$$E = \hbar\omega\left(n_x + n_y + n_z + \frac{3}{2}\right) = \hbar\omega\left(n + \frac{3}{2}\right)$$

The degeneracy is $$\frac{(n+1)(n+2)}{2}$$

The energy level structure is shown in the Table.

$E/\hbar\omega$	n_x	n_y	n_z	n	degeneracy
3/2	0	0	0	0	1
5/2	1	0	0	1	3
5/2	0	1	0	1	
5/2	0	0	1	1	
7/2	2	0	0	2	6
7/2	0	2	0	2	
7/2	0	0	2	2	
7/2	1	1	0	2	
7/2	1	0	1	2	
7/2	0	1	1	2	

Energy Levels - 3D Oscillator

In spherical-polar coordinates, we can follow a procedure similar to the plane- polar 2-dimensional case to get

$$\psi_{n_r,\ell,m}(y,\theta,\varphi) = y^{\ell} e^{-\frac{y^2}{2}} L_{n_r}^{\ell+\frac{1}{2}}(y^2) Y_{\ell m}(\theta,\varphi)$$

where

$$r = \rho y \quad , \quad \rho^2 = \frac{\hbar^2}{M\omega}$$

The corresponding energy values are

$$E_{n_r,\ell} = \hbar\omega(2n_r + \ell + \frac{3}{2})$$

which gives the Table.

$E/\hbar\omega$	n_r	ℓ	$n = 2n_r + \ell$	degeneracy
3/2	0	0	0	1
5/2	0	1	1	3 → m=± 1,0
7/2	1 or 0	0 or 2	2	6 → m=± 2, ± 1

Energy Levels - 3D Oscillator

The 3-Dimensional Delta Function

We now consider a particle is moving in 3-dimensions under the action of the attractive delta function potential given by

$$V(r) = -\frac{\hbar^2}{2M\alpha}\delta(r-a)$$

Since this is a central force we know that we can write the solutions in the form

$$\psi(r,\theta,\varphi) = R(r)Y_{\ell m}(\theta,\varphi)$$

The Schrodinger equation is

$$-\frac{\hbar^2}{2M}\left(\frac{d^2R}{dr^2}+\frac{2}{r}\frac{dR}{dr}-\frac{\ell(\ell+1)}{r^2}R\right)-\frac{\hbar^2}{2M\alpha}\delta(r-a)R-\frac{\hbar^2k^2}{2M}R=0$$

where, since we are looking for bound states, we have set

$$E = -\frac{\hbar^2k^2}{2M} < 0$$

Now the function $R(r)$ must be continuous everywhere.

But as we saw earlier in the 1-dimensional case, the derivative of R is not continuous for delta functions.

The discontinuity at $r = a$ is found by integrating the radial equation around the point $r = a$.

$$-\frac{\hbar^2}{2M}\int_{a^-}^{a^+} r^2dr\left(\frac{d^2R(r)}{dr^2}+\frac{2}{r}\frac{dR(r)}{dr}-\frac{\ell(\ell+1)}{r^2}R(r)\right)$$

$$-\frac{\hbar^2}{2M\alpha}\int_{a^-}^{a^+} r^2dr\delta(r-a)R(r)-\frac{\hbar^2k^2}{2M}\int_{a^-}^{a^+} r^2drR(r)=0$$

which gives the second boundary (continuity) condition

$$\alpha\left[\frac{dR(a^+)}{dr}-\frac{dR(a^-)}{dr}\right]=-R(a)$$

For $r>a$ and $r<a$ we then have the equation

$$-\frac{\hbar^2}{2M}\left(\frac{d^2R}{dr^2}+\frac{2}{r}\frac{dR}{dr}-\frac{\ell(\ell+1)}{r^2}R\right)-\frac{\hbar^2k^2}{2M}R=0$$

which has as a solution for the case $\ell=0$

$$R(r)=\frac{1}{r}(Ae^{-kr}+Be^{kr})$$

For $R(r)$ to be well behaved at $r=0$ we must choose $B=-A$.

Therefore, the solution for $r<a$ is

$$R(r)=\frac{c}{r}\sinh kr$$

For $R(r)$ to be well behaved as $r\to\infty$ we must choose $B=0$.

Therefore the solution for $r>a$ is

$$R(r)=\frac{b}{r}e^{-kr}$$

The boundary conditions at $r=a$ then give the equations

$$\frac{c}{a}\sinh ka=\frac{b}{a}e^{-ka}\rightarrow c\sinh ka=be^{-ka}$$

and

$$0=-k\alpha(be^{-ka}+c\cosh ka)+c\sinh ka$$

Eliminating b and c we get a transcendental equation for k (or E)

$$\frac{\alpha}{a} = \frac{1 - e^{-2ka}}{2ka}$$

The right-hand side of this equation has a range

$$0 < \frac{1 - e^{-2ka}}{2ka} < 1$$

Therefore, the allowed range of the parameter α, in order that an $\ell = 0$ bound state exist, is $0 < \alpha < a$.

The Hydrogen Atom

Schrodinger Equation Solution

The potential energy function $V(r) = -\frac{Ze^2}{r}$

represents the attractive Coulomb interaction between an atomic nucleus of charge $+Ze$ and an electron of charge $-e$.

The Schrodinger equation we have been using describes the motion of a single particle in an external field.

In the hydrogen atom, however, we are interested in the motion of two particles (nucleus and electron) that are attracted to each other via the potential above (where r is the separation distance between the two particles).

We start by writing the Schrodinger equation for the two particle system.

It involves six coordinates (three for each particle).

We use Cartesian coordinates to start the discussion.

We have

$$\hat{H} = \frac{\vec{p}_{1,op}^{\,2}}{2m_1} + \frac{\vec{p}_{2,op}^{\,2}}{2m_2} + V(\vec{r}) \quad where \quad \vec{r} = \vec{r}_1 - \vec{r}_2$$

which gives

$$-\frac{\hbar^2}{2m_1}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2}\right)\phi(x_1, y_1, z_1, x_2, y_2, z_2)$$
$$-\frac{\hbar^2}{2m_2}\left(\frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} + \frac{\partial^2}{\partial z_2^2}\right)\phi(x_1, y_1, z_1, x_2, y_2, z_2)$$
$$+ V(x_1 - x_2, y_1 - y_2, z_1 - z_2)\phi(x_1, y_1, z_1, x_2, y_2, z_2)$$
$$= E\phi(x_1, y_1, z_1, x_2, y_2, z_2)$$

We now introduce relative and center-of-mass(CM) coordinates by

$$x = x_1 - x_2 \quad , \quad y = y_1 - y_2 \quad , \quad z = z_1 - z_2$$
$$\vec{r} = (x, y, z)$$
$$MX = m_1 x_1 + m_2 x_2 \quad , \quad M\mathrm{Y} = m_1 y_1 + m_2 y_2 \quad , \quad M\mathrm{Z} = m_1 z_1 + m_2 z_2$$
$$\vec{R} = (X, Y, Z)$$
$$M = m_1 + m_2 = \text{ total mass of the system}$$

Substitution gives

$$-\frac{\hbar^2}{2M}\left(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2} + \frac{\partial^2}{\partial Z^2}\right)\phi(x, y, z, X, Y, Z)$$
$$-\frac{\hbar^2}{2\mu}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\phi(x, y, z, X, Y, Z)$$
$$+ V(x, y, z)\phi(x, y, z, X, Y, Z) = E\phi(x, y, z, X, Y, Z)$$

where

$$\mu = \frac{m_1 m_2}{m_1 + m_2} = \text{ the reduced mass}$$

We can now separate the variables by assuming a solution of the form

$$\phi(x, y, z, X, Y, Z) = \psi(x, y, z)\Psi(X, Y, Z)$$

which gives the two equations

$$-\frac{\hbar^2}{2\mu}\nabla^2_{\vec{r}}\psi(\vec{r}) + V(r)\psi(\vec{r}) = E\psi(\vec{r})$$

$$-\frac{\hbar^2}{2M}\nabla^2_{\vec{R}}\Psi(\vec{R}) = E'\Psi(\vec{R})$$

The second equation says that the CM of the two particles is like a free particle of mass M.

The first equation describes the relative motion of the two particles and is the same as the equation of motion of a particle of mass μ in an external potential energy $V(r)$.

In the hydrogen atom problem we are only interested in the energy levels E associated with the relative motion.

In addition, since the nuclear mass is so much larger than the electron mass, we have

$$\mu \approx m_e = m_{electron}$$

This is a central force so we assume a solution of the form

$$\psi(\vec{r}) = R(r)Y_{\ell m}(\theta, \varphi)$$

and obtain the radial equation

$$-\frac{\hbar^2}{2\mu}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{Ze^2}{r}R + \frac{\ell(\ell+1)\hbar^2}{2\mu r^2}R = ER$$

where $E < 0$ for a bound state.

We follow the same approach as before.

We change the variables so that the equation is in dimensionless form by introducing $\varrho = \alpha r$ where

$$\alpha^2 = \frac{8\mu|E|}{\hbar^2} \quad and \quad \lambda = \frac{2\mu Z e^2}{\alpha\hbar^2} = \frac{Ze^2}{\hbar}\left(\frac{\mu}{2|E|}\right)^{1/2}$$

We get

$$\frac{1}{\rho^2}\frac{d}{d\rho}\left(\rho^2\frac{dR}{dr}\rho\right) + \left(\frac{\lambda}{\rho} - \frac{1}{4} - \frac{\ell(\ell+1)}{\rho^2}\right)R = 0$$

For $\varrho \to \infty$ the equation becomes

$$\frac{1}{\rho^2}\frac{d}{d\rho}\left(\rho^2\frac{dR}{dr}\rho\right) - \frac{1}{4}R = 0$$

which has the solution

$$R \to \rho^n e^{\pm\frac{1}{2}\rho}$$

where n can have any finite value.

Since we want a normalizable solution, we will look for a solution of the form

$$R \to F(\rho)e^{-\frac{1}{2}\rho}$$

Substitution gives an equation for F

$$\frac{d^2F}{d\rho^2} + \left(\frac{2}{\rho} - 1\right)\frac{dF}{d\rho} + \left[\frac{\lambda - 1}{\rho} - \frac{\ell(\ell+1)}{\rho^2}\right]F = 0$$

We solve this equation (it is Laguerre's equation) by series substitution.

$$F(\rho) = \rho^s (a_0 + a_1\rho + a_2\rho^2 + \ldots.) = \rho^s L(\rho)$$

where $a_0 \neq 0$ and $s \geq 0$.

We must also have that $F(0)$ is finite.

Substitution gives an equation for L.

$$\rho^2 \frac{d^2L}{d\rho^2} + \rho\left[2(s+1) - \rho\right]\frac{dL}{d\rho} + \left[\rho(\lambda - s - 1) + s(s+1) - \ell(\ell+1)\right]L = 0$$

If we set $\varrho = 0$ in this equation and use the fact that L is a power series we get the condition

$$s(s+1) - \ell(\ell+1) = 0$$

or

$$s = \ell \text{ or } s = -(\ell+1)$$

Since R must be finite at the origin we exclude the second possibility and choose $s = \ell$.

This gives

$$\rho^2 \frac{d^2L}{d\rho^2} + \rho\left[2(\ell+1) - \rho\right]\frac{dL}{d\rho} + \left[\rho(\lambda - \ell - 1)\right]L = 0$$

Substituting the power series gives a recursion relation for the coefficients

$$a_{\nu+1} = \frac{\nu + \ell + 1 - \lambda}{(\nu+1)(\nu+2\ell+2)} a_\nu$$

If the series does not terminate, then this recursion relation behaves like

$$\frac{a_{\nu+1}}{a_\nu} \to \frac{1}{\nu}$$

for large ν.

This corresponds to the series for $\varrho^n e^{\varrho}$.

Since this will give a non-normalizable result, we must terminate the series by choosing

$$\lambda = n = \text{ positive integer } = n' + \ell + 1 = \text{ total quantum number}$$

where n' (= radial quantum number) is the largest power of ϱ in the solution L.

n' and n are non-negative integers, $n = 1,2,3,4,...$

We thus obtain the solution for the energies

$$E_n = -|E_n| = -\frac{\mu Z^2 e^4}{2\hbar^2 n^2} = -\frac{Z^2 e^2}{2a_0 n^2}$$

where

$$a_0 = \text{ Bohr radius } = \frac{\hbar^2}{\mu e^2}$$

Unlike the finite square well, where we had a finite number of bound state levels, in this case, we obtain an infinite set of discrete energies.

This results from the very slow decrease of the Coulomb potential energy with distance.

The Laguerre polynomial solutions are given by the generating function

$$G(\rho, s) = \frac{e^{-\frac{\rho s}{1-s}}}{1-s} = \sum_{q=0}^{\infty} \frac{L_q(\rho)}{q!} s^q \quad , \quad s < 1$$

By differentiating the generating function with respect to ϱ and s we can show that

$$\frac{dL_q}{d\rho} - q\frac{dL_{q-1}}{d\rho} = -qL_{q-1}$$

$$L_{q+1} = (2q + 1 - \rho)L_q - q^2 L_{q-1}$$

The lowest order differential equation involving only L_q that can be constructed from these two equations is

$$\rho\frac{d^2L_q}{d\rho^2}+(1-\rho)\frac{dL_q}{d\rho}+qL_q=0$$

This is not quite our original equation.

However, if we define the associated Laguerre polynomials by

$$L_q^p(\rho)=\frac{d^p}{d\rho^p}L_q(\rho)$$

then differentiating the differential equation p times we get

$$\rho\frac{d^2L_q^p}{d\rho^2}+(p+1-\rho)\frac{dL_q^p}{d\rho}+(q-p)L_q^p=0$$

Setting $\lambda = n$ we then see that the solutions to the Schrodinger equation are

$$L_{n+\ell}^{2\ell+1}(\rho)$$

which are polynomials of order $(n+\ell)-(2\ell+1)=n-\ell-1$ in agreement with the earlier results.

We can differentiate the generating function p times to get

$$G_p(\rho,s)=\frac{(-s)^p e^{-\frac{\rho s}{1-s}}}{(1-s)^{p+1}}=\sum_{q=0}^{\infty}\frac{L_q^p(\rho)}{q!}s^q \quad , \quad s<1$$

Explicitly, we then have

$$L_{n+\ell}^{2\ell+1}(\rho)=\sum_{k=0}^{n-\ell-1}(-1)^{k+1}\frac{[(n+\ell)!]^2\rho^k}{(n-\ell-1-k)!(2\ell+1+k)!k!}$$

The normalized radial wave functions are of the form

$$R_{n\ell}(\rho) = -\left\{\left(\frac{2Z}{na_0}\right)^3 \frac{(n-\ell-1)!}{2n\left[(n+\ell)!\right]^3}\right\}^{1/2} e^{-\frac{1}{2}\rho}\rho^{\ell} L_{n+\ell}^{2\ell+1}(\rho)$$

with

$$a_0 = \frac{\hbar^2}{\mu e^2} \text{ and } \rho = \frac{2Z}{na_0} r$$

and

$$\psi_{n\ell m}(r,\theta,\varphi) = R_{n\ell}(r) Y_{\ell m}(\theta,\varphi)$$

The first few radial wave functions are

$$R_{10}(r) = \left(\frac{Z}{a_0}\right)^{3/2} 2e^{-\frac{Zr}{a_0}}$$

$$R_{20}(r) = \left(\frac{Z}{2a_0}\right)^{3/2} \left(2 - \frac{Zr}{a_0}\right) e^{-\frac{Zr}{2a_0}}$$

$$R_{21}(r) = \left(\frac{Z}{2a_0}\right)^{3/2} \frac{Zr}{a_0\sqrt{3}} e^{-\frac{Zr}{2a_0}}$$

What about degeneracy since the energy values do not depend on ℓ and m?

For each value of n, ℓ can vary between 0 and $n-1$, and for each value of these ℓ values m can vary between $-\ell$ and ℓ ($2\ell + 1$ values).

Therefore the total degeneracy of an energy level E_n is given by

$$\sum_{\ell=0}^{n-1} (2\ell+1) = 2\frac{n(n-1)}{2} + n = n^2$$

The degeneracy with respect to *m* is true for any central force (as we have seen in other examples).

The ℓ degeneracy, however, is characteristic of the Coulomb potential alone.

It is called an *accidental degeneracy*.

Some useful expectation values are:

$$\langle r\rangle_{n\ell m} = \frac{a_0}{2Z}\left[3n^2 - \ell(\ell+1)\right] \qquad \left\langle \frac{1}{r}\right\rangle_{n\ell m} = \frac{Z}{a_0 n^2}$$

$$\left\langle \frac{1}{r^2}\right\rangle_{n\ell m} = \frac{Z^2}{a_0^2 n^3(\ell+\frac{1}{2})} \qquad \left\langle \frac{1}{r^3}\right\rangle_{n\ell m} = \frac{Z^3}{a_0^3 n^3 \ell(\ell+\frac{1}{2})(\ell+1)}$$

Algebraic Solution of the Hydrogen Atom

Review of the Classical Kepler Problem

In the classical Kepler problem of an electron of charge -*e* moving in the electric force field of a positive charge *Ze* located at the origin, the angular momentum

$$\vec{L} = \vec{r} \times \vec{p}$$

is constant.

The motion takes place in a plane perpendicular to the constant angular momentum direction.

Newton's second law tells us that the rate of change of the momentum is equal to the force

$$\frac{d\vec{p}}{dt} = -\frac{Ze^2}{r^2}\hat{r} \text{ where } r = |\vec{r}| \text{ and } \hat{r} = \frac{\vec{r}}{r} = \text{ unit vector}$$

The so-called Runge-Lenz vector is defined by

$$\vec{A} = \frac{1}{Ze^2 m}(\vec{L} \times \vec{p}) + \hat{r}$$

If we take its time derivative we find

$$\frac{d\vec{L}}{dt} = 0 \text{ and } \vec{L} = m\vec{r} \times \frac{d\vec{r}}{dt} = mr^2\hat{r} \times \frac{d\hat{r}}{dt} + mr(\hat{r} \times \hat{r})\frac{dr}{dt} = mr^2\hat{r} \times \frac{d\hat{r}}{dt}$$

$$\frac{d\vec{A}}{dt} = \frac{1}{Ze^2m}(\vec{L} \times \frac{d\vec{p}}{dt} + \frac{d\vec{L}}{dt} \times \vec{p}) + \frac{d\hat{r}}{dt} = \frac{1}{Ze^2m}(\vec{L} \times \frac{d\vec{p}}{dt}) + \frac{d\hat{r}}{dt}$$

$$= -\frac{1}{m}(\vec{L} \times \frac{\hat{r}}{r^2}) + \frac{d\hat{r}}{dt} = -(\hat{r} \times \frac{d\hat{r}}{dt}) \times \hat{r} + \frac{d\hat{r}}{dt}$$

$$= -(\frac{d\hat{r}}{dt}(\hat{r} \cdot \hat{r}) - (\hat{r} \cdot \frac{d\hat{r}}{dt})) + \frac{d\hat{r}}{dt} = -\frac{d\hat{r}}{dt} + \frac{d(\hat{r} \cdot \hat{r})}{dt} + \frac{d\hat{r}}{dt} = 0$$

Thus, the vector $\vec{A}$ is a constant of the motion.

It corresponds physically to the length and direction of the semi-major axis of the classical elliptical orbit.

The equation of the elliptical orbit is easily found using the $\vec{A}$ vector.

$$\vec{A} \cdot \vec{r} = ar\cos\theta = \frac{1}{Ze^2m}(\vec{L} \times \vec{p}) \cdot \vec{r} + \hat{r} \cdot \vec{r}$$

$$= -\frac{1}{Ze^2m}\vec{L} \cdot (\vec{r} \times \vec{p}) + r$$

$$ar\cos\theta = -\frac{L^2}{Ze^2m} + r$$

$$\rightarrow \frac{1}{\mathrm{r}} = \frac{Ze^2m}{L^2}(1 - a\cos\theta) \rightarrow \text{ orbit equation (conic sections)}$$

where

$$a = |\vec{A}|$$

= eccentricity and the direction of $\vec{A}$ is from the origin to the aphelion

The Quantum Mechanical Problem

In order to make this useful in quantum mechanics, the classical Runge-Lenz vector must be generalized.

In classical physics

$$\vec{L} \times \vec{p} = -\vec{p} \times \vec{L}$$

In quantum mechanics, this identity is not valid since the components of $\vec{L}_{op}$ do not commute.

The correct quantum mechanical generalization (remember it must be a Hermitian operator if it is a physical observable) of the vector $\vec{A}$ is

$$\vec{A}_{op} = \frac{1}{2Ze^2m}(\vec{L}_{op} \times \vec{p}_{op} - \vec{p}_{op} \times \vec{L}_{op}) + \left(\frac{\vec{r}}{r}\right)_{op}$$

This satisfies $[\hat{H}, \vec{A}_{op}] = 0$, which says $\vec{A}$ is a constant.

It also satisfies $\vec{A}_{op} \cdot \vec{L}_{op} = 0$.

We can derive the following commutators:

$$[\hat{L}_i, \hat{A}_j] = i\hbar\varepsilon_{ijk}\hat{A}_k \quad , \quad [\hat{L}_i, \hat{p}_j] = i\hbar\varepsilon_{ijk}\hat{p}_k \quad , \quad [\hat{L}_i, \hat{r}_j] = i\hbar\varepsilon_{ijk}\hat{r}_k$$

which gives

$$\vec{A}_{op} = \frac{1}{2Ze^2m}(2\vec{L}_{op} \times \vec{p}_{op} + i\hbar\vec{p}_{op}) + \left(\frac{\vec{r}}{r}\right)_{op}$$

We now derive two important properties of $\vec{A}$.

The first property follows from the commutators

$$\left[\left(\vec{L}_{op}\times\vec{p}_{op}\right)_i,\hat{p}_j\right]+\left[\hat{p}_i,\left(\vec{L}_{op}\times\vec{p}_{op}\right)_j\right]=0$$

$$\left[\left(\vec{L}_{op}\times\vec{p}_{op}\right)_i,\left(\vec{L}_{op}\times\vec{p}_{op}\right)_j\right]=-i\hbar\varepsilon_{ijk}\hat{L}_k p^2=-i\hbar\varepsilon_{ijk}p^2\hat{L}_k$$

$$\left[\left(\vec{L}_{op}\times\vec{p}_{op}\right)_i,\frac{\hat{r}_j}{r}\right]+\left[\frac{\hat{r}_i}{r},\left(\vec{L}_{op}\times\vec{p}_{op}\right)_j\right]=2i\hbar\varepsilon_{ijk}\frac{\hat{L}_k}{r}$$

which leads to

$$\left[\hat{A}_i,\hat{A}_j\right]=i\hbar\left(\frac{-2\hat{H}}{Z^2e^4m}\right)\varepsilon_{ijk}\hat{L}_k$$

where

$$\hat{H}=\text{ Hamiltonian }=\frac{\vec{p}_{op}^{\,2}}{2m}-Ze^2\left(\frac{1}{r}\right)_{op}$$

The second property follows from the relations

$$(\vec{L}_{op}\times\vec{p}_{op})\cdot(\vec{L}_{op}\times\vec{p}_{op})=\vec{L}_{op}^2\vec{p}_{op}^2$$

$$\left(\frac{\vec{r}}{r}\right)_{op}\cdot(\vec{L}_{op}\times\vec{p}_{op})+(\vec{L}_{op}\times\vec{p}_{op})\cdot\left(\frac{\vec{r}}{r}\right)_{op}=-\frac{2\vec{L}_{op}^2}{r}+2i\hbar\left(\frac{\vec{r}}{r}\right)_{op}\cdot\vec{p}_{op}$$

$$\vec{p}_{op}\cdot\left(\frac{\vec{r}}{r}\right)_{op}=\left(\frac{\vec{r}}{r}\right)_{op}\cdot\vec{p}_{op}-2i\hbar\left(\frac{1}{r}\right)_{op}$$

$$\vec{p}_{op}\cdot\vec{L}_{op}=\vec{L}_{op}\cdot\vec{p}_{op}=0$$

$$\vec{p}_{op}\cdot(\vec{L}_{op}\times\vec{p}_{op})+(\vec{L}_{op}\times\vec{p}_{op})\cdot\vec{L}_{op}=2i\hbar\vec{p}_{op}^2$$

which lead to

$$\vec{A}_{op}^2=\vec{A}_{op}\cdot\vec{A}_{op}=1+\frac{2\hat{H}}{Z^2e^4m}(\vec{L}_{op}^2+\hbar^2)$$

We now define the two new operators (they are *ladder operators*) by

$$\vec{I}_{op}^{\pm} = \frac{1}{2}\left[\vec{L}_{op} \pm \left(-\frac{2\hat{H}}{Z^2 e^4 m}\right)^{1/2} \vec{A}_{op}\right]$$

Using the relations we have derived for $\left[\hat{A}_i, \hat{A}_j\right]$ and $\left[\hat{L}_i, \hat{A}_j\right]$ we have

$$\left[\vec{I}_{op}^{+}, \vec{I}_{op}^{-}\right] = 0$$

which means they have a common eigenbasis.

We also have

$$\left[\hat{I}_{op,i}^{\pm}, \hat{I}_{op,j}^{\pm}\right] = i\hbar\varepsilon_{ijk}\hat{I}_{op,k}^{\pm}$$

which are the standard angular momentum component commutation relations.

Since they each commute with $\hat{H}$, they also have a common eigenbasis with $\hat{H}$.

Therefore, we can find a set of states such that

$$(\vec{I}_{op}^{\pm})^2 |\psi\rangle = i_{\pm}(i_{\pm}+1)\hbar^2 |\psi\rangle \text{ and } \hat{H}|\psi\rangle = E|\psi\rangle$$

We can show that $(\vec{I}_{op}^{+})^2 = (\vec{I}_{op}^{-})^2$, which implies that $i_+ = i_-$.

We also have(as before)

$$i_+ = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3, \ldots$$

Since $\vec{A}_{op} \cdot \vec{L}_{op} = 0$ we get

$$2\left[(\vec{I}_{op}^{+})^2 + (\vec{I}_{op}^{-})^2\right] + \hbar^2 = -\frac{Z^2 e^4 m}{2\hat{H}}$$

$$\left[4i_+(i_+ + 1) + 1\right]\hbar^2 = -\frac{Z^2 e^4 m}{2E}$$

$$E = -\frac{Z^2 e^4 m}{2(2i_+ + 1)^2} = -\frac{Z^2 e^4 m}{2n^2}$$

where we have set $n = 2i_+ + 1 = 1,2,3,4,5,...$, which are the correct energy values for hydrogen.

While the energy depends only on the quantum number i_+ (or n) each state has a degeneracy depending on the number of z-component values for each i_+ value.

This is

$$(2i_+ + 1)(2i_- + 1) = (2i_+ + 1)^2 = n^2$$

which is the correct degeneracy.

The Deuteron

A deuteron is a bound state of a neutron and a proton.

We can consider this system as a single particle with reduced mass

$$\mu = \frac{m_{proton} m_{neutron}}{m_{proton} + m_{neutron}} \approx \frac{m_{proton}}{2} = \frac{m_p}{2}$$

moving in a fixed potential $V(r)$, where r is the proton-neutron separation.

As a first approximation, we assume that the nuclear interaction binding the proton and the neutron into a deuteron is a finite square well in 3–dimensions.

$$V(r) = \begin{cases} -V_0 & r < a \\ 0 & r \geq a \end{cases}$$

The physical properties of the deuteron system are:

1. Almost an $\ell = 0$ state (a small admixture of $\ell = 2$ is present). We will assume $\ell = 0$.

2. Only one bound state exists.

3. The depth and range of the potential is such that the deuteron is weakly bound. The energy level in the potential corresponds to the binding energy of the system, where

$$E = \text{ binding energy } = m_{deuteron}c^2 - (m_{proton} + m_{neutron})c^2 < 0 \quad (7.683)$$

Experimentally, it has been found that $E = -2.228\,MeV$.

4. By weakly bound we mean

$$\frac{|E|}{(m_{proton} + m_{neutron})c^2} << 1 \quad (7.684)$$

In fact, for the deuteron we have

$$\frac{|E|}{(m_{proton} + m_{neutron})c^2} \approx 0.001 \quad (7.685)$$

5. This system is so weakly bound that any small decrease in the radius of the well a or small reduction in the depth of the well V_0 would cause the system to break up (no bound state exists).

We derived the solutions for this potential earlier.

$$R(r) = Aj_0(\alpha r) \qquad r < a$$

$$R(r) = Bh_0^{(1)}(i\beta r) \qquad r \geq a$$

where

$$\alpha^2 = \frac{2m}{\hbar^2}(V_0 - |E|) \text{ and } \beta^2 = \frac{2m}{\hbar^2}|E|$$

The transcendental equation arising from matching boundary conditions at $r = a$ gave us the equations

$$\eta = -\xi \cot \xi \text{ and } \xi^2 + \eta^2 = \frac{2mV_0a^2}{\hbar^2}$$

where

$$\xi = \alpha a \text{ and } \eta = \beta a$$

The graphical solution as shown below in the Figure plots

$$\eta = -\xi \cot \xi \text{ and } \eta^2 = \frac{2mV_0a^2}{\hbar^2} - \xi^2 \text{ versus } \xi$$

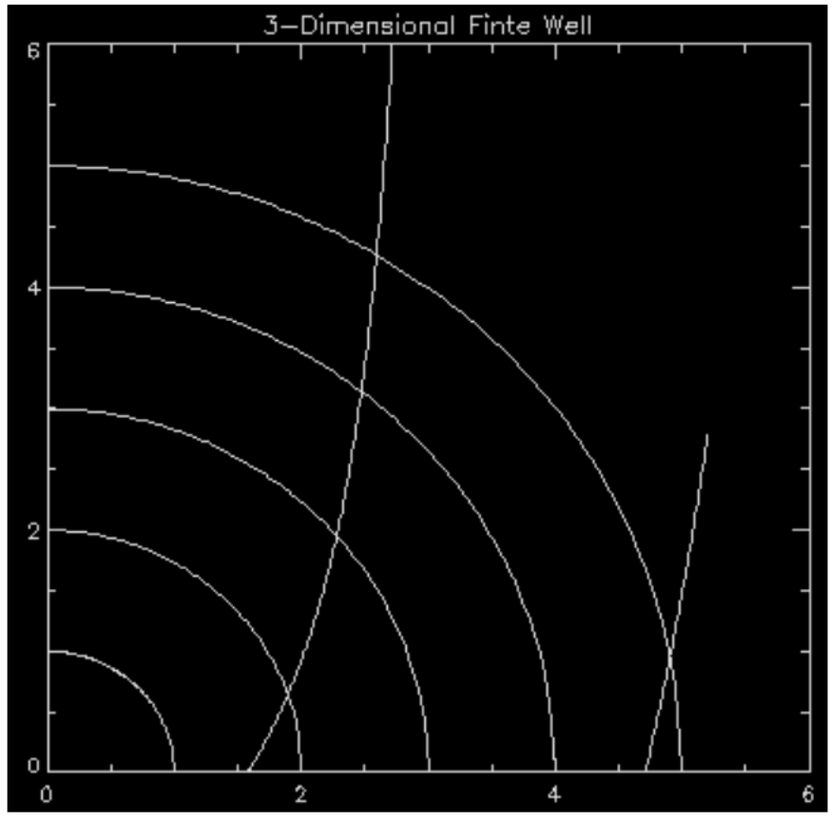
We found a finite number of bound states for given values of the well parameters.

In particular

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{\pi}{2}\right)^2 \rightarrow \text{ no solution}$$

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{3\pi}{2}\right)^2 \rightarrow \text{ 1 solution}$$

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{5\pi}{2}\right)^2 \rightarrow \text{ 2 solutions}$$



Deuteron Solution

For the weakly bound deuteron system we expect that

$$\xi = \frac{\pi}{2} + \varepsilon \quad , \quad \varepsilon \ll \frac{\pi}{2}$$

i.e., the radius of the circle is barely large enough to get a single (weakly) bound state.
Substituting into the transcendental equation we get

$$\eta = -\left(\frac{\pi}{2}+\varepsilon\right)\cot\left(\frac{\pi}{2}+\varepsilon\right) = -\left(\frac{\pi}{2}+\varepsilon\right)\frac{\cos\left(\frac{\pi}{2}+\varepsilon\right)}{\sin\left(\frac{\pi}{2}+\varepsilon\right)}$$

$$= -\left(\frac{\pi}{2}+\varepsilon\right)\frac{\cos\left(\frac{\pi}{2}\right)\cos(\varepsilon) - \sin\left(\frac{\pi}{2}\right)\sin(\varepsilon)}{\sin\left(\frac{\pi}{2}\right)\cos(\varepsilon) + \cos\left(\frac{\pi}{2}\right)\sin(\varepsilon)}$$

$$\approx \left(\frac{\pi}{2}+\varepsilon\right)\frac{\sin(\varepsilon)}{\cos(\varepsilon)} \approx \left(\frac{\pi}{2}+\varepsilon\right)\frac{\varepsilon - \frac{\varepsilon^3}{6}}{1-\frac{\varepsilon^2}{2}} \approx \left(\frac{\pi}{2}+\varepsilon\right)\left(\varepsilon - \frac{\varepsilon^3}{6}\right)\left(1+\frac{\varepsilon^2}{2}\right)$$

$$\approx \frac{\pi}{2}\varepsilon + (1+\frac{\pi}{4})\varepsilon^2$$

Now substituting into the circle equation we get

$$\left(\frac{\pi}{2}+\varepsilon\right)^2 + \left(\frac{\pi}{2}\varepsilon + (1+\frac{\pi}{4})\varepsilon^2\right)^2 = \frac{2mV_0a^2}{\hbar^2}$$

$$\frac{\pi^2}{4} + \pi\varepsilon + \left(1+\frac{\pi^2}{4}\right)\varepsilon^2 \approx \frac{2mV_0a^2}{\hbar^2}$$

Dropping the small quadratic terms in ε we get

$$\varepsilon \approx \frac{2mV_0a^2}{\pi\hbar^2} - \frac{\pi}{4}$$

Therefore,

$$|E| = \frac{\hbar^2}{2ma^2}\eta^2 \approx \frac{\hbar^2\pi^2}{8ma^2}\left(\frac{2mV_0a^2}{\pi\hbar^2} - \frac{\pi}{4}\right)^2$$

A typical value for the range of the interaction is the order of 2 Fermi or $a \approx 2\times10^{-13}$ cm.

In order to get $|E| \approx 2.3$ MeV, we would need a well depth of $V_0 \approx 42$ M eV which is reasonable (according to experimentalists).

The Deuteron - Another Way

Experiment (scattering) indicates that instead of a square well (very unrealistic) the actual potential is of the form

$$V(r) = -Ae^{-\frac{r}{a}}$$

where

$$A \approx 32MeV \text{ and } a \approx 2.25x10^{-13}cm$$

We can solve for this potential ($\ell = 0$ case) exactly using this clever trick I first learned from Hans Bethe at Cornell University.

The radial equation in this case is

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + \frac{2m}{\hbar^2}\left(Ae^{-\frac{r}{a}} - |E|\right)R = 0$$

$$\frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} + \frac{2m}{\hbar^2}\left(Ae^{-\frac{r}{a}} - |E|\right)R = 0$$

Now let

$$R(r) = \frac{\chi(r)}{r}$$

which implies that

$$\frac{dR}{dr} = \frac{d\left(\frac{\chi}{r}\right)}{dr} = \frac{1}{r}\frac{d\chi}{dr} - \frac{1}{r^2}\chi$$
$$\frac{d^2R}{dr^2} = -\frac{2}{r^2}\frac{d\chi}{dr} + \frac{1}{r}\frac{d^2\chi}{dr^2} + \frac{2}{r^3}\chi$$

Substitution gives

$$\frac{d^2\chi}{dr^2} + \frac{2m}{\hbar^2}\left(Ae^{-\frac{r}{a}} - |E|\right)\chi = 0$$

We now change the variables using

$$\xi = e^{-\frac{r}{2a}} \rightarrow \frac{d}{dr} = \frac{d\xi}{dr}\frac{d}{d\xi} = -\frac{\xi}{2a}\frac{d}{d\xi}$$
$$\frac{d^2}{dr^2} = \frac{d}{dr}\left(-\frac{\xi}{2a}\frac{d}{d\xi}\right) = \frac{d\xi}{dr}\frac{d}{d\xi}\left(-\frac{\xi}{2a}\frac{d}{d\xi}\right) = \left(\frac{\xi}{2a}\right)^2\frac{d^2}{d\xi^2} + \frac{\xi}{4a^2}\frac{d}{d\xi}$$

We then get the equation

$$\xi^2\frac{d^2\chi}{d\xi^2} + \xi\frac{d\chi}{d\xi} + ((\alpha a)^2\xi^2 - (ka)^2)\chi = 0$$

where

$$(\alpha a)^2 = \frac{2mA}{\hbar^2}a^2 \text{ and } (ka)^2 = \frac{2m|E|}{\hbar^2}a^2$$

Now Bessel's equation has the form

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - \nu^2)y = 0$$

Therefore, we have Bessel's equation with a general solution

$$\chi(r) = CJ_{ka}(\alpha a\xi) + BY_{ka}(\alpha a\xi)$$

$$R(r) = \frac{1}{r}(CJ_{ka}(\alpha a\xi) + BY_{ka}(\alpha a\xi))$$

As $r \to \infty$, $\xi = e^{-r} \to 0$, which implies we must choose **B** = 0 or the solution diverges.

As $r \to 0$, $\xi = e^{-r} \to 1$, which implies we must have

$$J_{ka}(\alpha a) = 0$$

or the solution diverges.

This is then the energy eigenvalue condition.

For the values of *A* and a given earlier we have

$$J_{ka}(6.28) = 0$$

and if we let $|E| = qA$, this becomes

$$J_{1/2}(6.28) = 0$$

Now

$$J_{6.4q}(6.28) = 0$$

Therefore we have

$$6.4q = \frac{1}{2} \to q = \frac{1}{12.8} \to |E| = qA = 2.34\,MeV$$

which is an excellent result for the bound state energy of the deuteron.

Linear Potential

We now consider a *linear* potential energy function given by

$$V(r) = \alpha r - V_0$$

The Schrodinger equation becomes

$$-\frac{\hbar^2}{2m}\left(\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right) - \frac{\vec{L}_{op}^2}{\hbar^2 r^2}\psi\right) + (\alpha r - V_0 - E)\psi = 0$$

Since it is a central potential, we can write

$$\psi = R(r)Y_{\ell m}(\theta, \varphi)$$

which implies

$$-\frac{\hbar^2}{2m}\left(\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{\ell(\ell+1)}{r^2}R\right) + (\alpha r - V_0 - E)R = 0$$

If we let

$$\chi(r) = rR(r)$$

we get

$$-\frac{\hbar^2}{2m}\frac{d^2\chi}{dr^2} + (\alpha r - V_0 - E + \frac{\hbar^2\ell(\ell+1)}{2mr^2})\chi = 0$$

For $\ell > 0$ there is no closed form solution for this equation and numerical methods must be applied.

For $\ell = 0$, we have

$$-\frac{\hbar^2}{2m}\frac{d^2\chi}{dr^2} + (\alpha r - V_0 - E)\chi = 0$$

Now we let

$$\xi = \left(r - \frac{E + V_0}{\alpha}\right)\left(\frac{2m\alpha}{\hbar^2}\right)^{1/3}$$

which implies that

$$\frac{d}{dr} = \frac{d\xi}{dr}\frac{d}{d\xi} = \left(\frac{2m\alpha}{\hbar^2}\right)^{1/3}\frac{d}{d\xi}$$

$$\frac{d^2}{dr^2} = \frac{d}{dr}\left(\frac{d\xi}{dr}\frac{d}{d\xi}\right) = \frac{d\xi}{dr}\frac{d}{d\xi}\left(\left(\frac{2m\alpha}{\hbar^2}\right)^{1/3}\frac{d}{d\xi}\right)$$

$$= \left(\frac{2m\alpha}{\hbar^2}\right)^{2/3}\frac{d^2}{d\xi^2}$$

We then get the equation

$$\frac{d^2\chi}{d\xi^2} - \xi\chi = 0$$

This equation is *not as simple* as it looks.

Let us try a series solution of the form

$$\chi(\xi) = \sum_{n=0}^{\infty} a_n \xi^n$$

Substitution gives the relations

$$a_2 = 0 \text{ and } a_{m+2} = \frac{a_{m-1}}{(m+2)(m+1)}$$

The solution that goes to zero as $\xi \to \pm\infty$ is then of the form

$$\chi(\xi) = c_1 f(\xi) - c_2 g(\xi) = CAi(\xi) = \text{ Airy function}$$

where

$$f(x) = 1 + \frac{1}{3!}x^3 + \frac{1\cdot 4}{6!}x^6 + \frac{1\cdot 4\cdot 7}{9!}x^9 + \ldots$$

$$g(x) = x + \frac{2}{4!}x^4 + \frac{2\cdot 5}{7!}x^7 + \frac{2\cdot 5\cdot 8}{10!}x^{10} + \ldots$$

Now to insure that the wave function is normalizable we must also have

$$\chi(\xi(r=0))=0$$

$$Ai\left(-\frac{E+V_0}{\alpha}\left(\frac{2m\alpha}{\hbar^2}\right)^{1/3}\right)=0$$

This says that

$$-\frac{E_n+V_0}{\alpha}\left(\frac{2m\alpha}{\hbar^2}\right)^{1/3}=z_n=n^{th} \text{ zero of } Ai(\xi)$$

Thus, the allowed energies are given by

$$E_n=-z_n\left(\frac{\hbar^2\alpha^2}{2m}\right)^{1/3}-V_0$$

The first three zeroes are

$$z_1=-2.3381 \quad , \quad z_2=-4.0879 \quad , \quad z_3=-5.5209$$

Programming the numerical method we described in the book for solving the Schrodinger equation allows us to determine *E* values where for simplicity, we have chosen $\hbar = m = \alpha = V_0 = 1$ which gives the $\ell = 0$ energies as

$$E_n=-z_n\left(\frac{1}{2}\right)^{1/3}-1=-0.794z_n-1$$

The general ℓ equation is (in this case)

$$\frac{d^2\chi}{dr^2}-2(r-1-E+\frac{\ell(\ell+1)}{2r^2})\chi=0$$

For $\ell = 0$ we theoretically expect the first three energy values 0.856, 2.246 and 3.384 and the program produces the values 0.855750, 2.24460, 3.38160 which is good agreement.

We now use the linear potential to look at quark-quark bound states at low energies.

Modified Linear Potential and Quark Bound States

Over the past several decades the quark model of elementary particles has had many successes.

Some experimental results are the following:

1. Free (isolated) quarks have never been observed.
2. At small quark separations *color charge* exhibits behavior similar to that of ordinary charge.
3. Quark-Antiquark pairs form bound states.

We can explain some of the features of these experimental results by describing the quark-quark force by an effective potential of the form

$$V(r) = -\frac{A}{r} + Br$$

and using non-relativistic quantum mechanics.

The linear term corresponds to a long-range confining potential that is responsible for the fact that free, isolated quarks have not been observed.

In simple terms, a linear potential of this type means that it costs more and more energy as one of the quarks in a bound system attempts to separate from the other quark.

When this extra energy is twice the rest energy of a quark, a new quark-antiquark pair can be produced.

So instead of the quark getting free, one of the newly created quarks joins with one of the original quarks to recreate the bound pair (so it looks like nothing has happened) and the other new quark binds with the quark attempting to get free into a new meson.

We never see a free quark!

A lot of energy has been expended, but the outcome is the creation of a meson rather than the appearance of free quarks.

The other term, which resembles a Coulomb potential, reflects the fact that at small separations the so-called "color charge" forces behave like ordinary charge forces.

The original observations of quark-quark bound states was in 1974.

The experiments involved bound states of charmed quarks called charmonium (named after the similar bound states of electrons and positrons called positronium).

The observed bound-state energy levels were as shown in the Table:

n	ℓ	$E(GeV)$
1	0	3.097
2	1	3.492
2	0	3.686
3	0	4.105
4	0	4.414

Observed Bound-State Energy Levels

The Schrodinger equation for this potential becomes

$$-\frac{\hbar^2}{2m}\left(\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right)-\frac{\vec{L}_{op}^2}{\hbar^2 r^2}\psi\right)+(V(r)-E)\psi=0$$

Since it is a central potential, we can write

$$\psi = R(r)Y_{\ell m}(\theta,\varphi)$$

which implies

$$-\frac{\hbar^2}{2m}\left(\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)-\frac{\ell(\ell+1)}{r^2}R\right)+(V(r)-E)R=0$$

If we let

$$\chi(r)=rR(r)$$

we get

$$-\frac{\hbar^2}{2m}\frac{d^2\chi}{dr^2}+(V(r)-E+\frac{\hbar^2\ell(\ell+1)}{2mr^2})\chi=0$$

or

$$\frac{d^2\chi}{dr^2}-(aE-\frac{\ell(\ell+1)}{r^2}+\frac{b}{r}-cr)\chi=0$$

We must solve this system numerically.

The same program as earlier works again with a modified potential function.

The results for a set of parameters (a = 0.0385, b = 2.026, c = 34.65) chosen to get the right relationship between the levels are shown in the Table:

n	ℓ	$E(calculated)$	$E(rescaled)$
1	0	656	3.1
2	1	838	3.4
2	0	1160	3.6
3	0	1568	4.1
4	0	1916	4.4

Quark Model-Numerical Results

which is a reasonably good result and indicates that validity of the model.

The rescaled values are adjusted to correspond to the earlier table.

A more exact parameter search produces almost exact agreement.

Chapter 8

Time-Independent Perturbation Theory

Nondegenerate Case

For most physically interesting systems, it is not possible to find simple, exact formulas for the energy eigenvalues and state vectors.

In many cases, however, the real system is very similar to another system that we can solve exactly in closed form.

Our procedure will then be to approximate the real system by the similar system and approximately calculate corrections to find the corresponding values for the real system.

The approximation method that is most often used is called perturbation theory.

Rayleigh-Schrodinger Perturbation Theory

Consider the problem of finding the energies (eigenvalues) and state vectors (eigenvectors) for a system with a Hamiltonian $\hat{H}$ that can be written in the form

$$\hat{H} = \hat{H}_0 + \hat{V}$$

where we have already solved the system described by $\hat{H}_0$, i.e., we know that

$$\hat{H}_0 |n\rangle = \varepsilon_n |n\rangle$$

with $\langle m|n\rangle = \delta_{mn}$ (remember that the eigenvectors of a Hermitian operator always form a complete orthonormal set ... or we can make them so using the Gram-Schmidt process if degeneracy exists).

We call this solvable system the *unperturbed* or *zero-order* system.

We then assume that the extra term $\hat{V}$ is a *small correction* to $\hat{H}_0$ (that is what we mean by *similar* systems).

This says that the real physical system will have a solution given by

$$\hat{H}\,|N\rangle = E_n\,|N\rangle$$

where the real physical state vectors $|N\rangle$ are only slightly different from the unperturbed state vectors $|n\rangle$ and the real physical energies E_n are only slightly different from the unperturbed energies ε_n.

Mathematically, we can express this situation by writing the perturbation in the form

$$\hat{V} = g\hat{U}$$

where g is some small ($\ll 1$) constant factor pulled out of the correction term $\hat{V}$ that characterizes its strength (or effect on the system described by $\hat{H}_0$) of the perturbation.

As $g \to 0$, each eigenvector $|N\rangle$ of $\hat{H}$ must approach the corresponding eigenvector $|n\rangle$ of $\hat{H}_0$ and each energy eigenvalue E_n of $\hat{H}$ must approach the corresponding energy eigenvalue ε_n of $\hat{H}_0$.

We can guarantee that this property is true by assuming that power series exin the small parameter g exist for all physically relevant quantities of the real system, i.e..

$$\hat{H} = \hat{H}_0 + \hat{V} = \hat{H}_0 + g\hat{U}$$

$$|N\rangle = |n\rangle + g\left|N^{(1)}\right\rangle + g^2\left|N^{(2)}\right\rangle + \ldots$$

$$E_n = \varepsilon_n + gE_n^{(1)} + g^2E_n^{(2)} + \ldots$$

where the terms $|N^{(i)}\rangle$ and $E^{(i)}$ are called the i^{th}-order correction to the unperturbed or zero-order solution.

This is a major assumption, that we cannot, in general, prove is true a priori, i.e., we cannot prove that the power series converge and therefore make sense.

The usual normalization condition we might impose would be $\langle N \mid N \rangle = 1$.

Since the results of any calculation are independent of the choice of normalization (remember the expectation value and density operator definitions all include the norm in the denominator), we choose instead to use the normalization condition

$$\langle n \mid N \rangle = 1$$

which will greatly simplify our derivations and subsequent calculations.

Substituting the power series expansion into the normalization condition we get

$$\langle n \mid N \rangle = 1 = \langle n \mid n \rangle + g \left\langle n \mid N^{(1)} \right\rangle + g^2 \left\langle n \mid N^{(2)} \right\rangle + \ldots$$

But since we already have assumed that $\langle n \mid n \rangle = 1$, we must have

$$0 = g \left\langle n \mid N^{(1)} \right\rangle + g^2 \left\langle n \mid N^{(2)} \right\rangle + \ldots.$$

Now the only way for a power series to be *identically* zero is for the coefficient of each power of g to be separately equal to zero.

This gives the result

$$\left\langle n \mid N^{(i)} \right\rangle = 0 \quad , \quad i = 1, 2, 3, 4, \ldots$$

as a direct consequence of the normalization condition, i.e., all corrections to the state vector are orthogonal to the unperturbed state vector.

We now substitute all of these power series into the original energy eigenvalue equation for $\hat{H}$:

$$\hat{H}\,|N\rangle = E_n\,|N\rangle$$

$$\left(\hat{H}_0 + g\hat{U}\right)\left(|n\rangle + g\left|N^{(1)}\right\rangle + g^2 + \left|N^{(2)}\right\rangle ....\right)$$

$$= \left(\varepsilon_n + gE_n^{(1)} + g^2E_n^{(2)} + ...\right)\left(|n\rangle + g\left|N^{(1)}\right\rangle + g^2 + \left|N^{(2)}\right\rangle ....\right)$$

We now multiply everything out and collect terms in a single power series in g.

We get

$$0 = \left(\hat{H}_0\,|n\rangle - \varepsilon_n\,|n\rangle\right)g^0 + \left(\hat{H}_0\left|N^{(1)}\right\rangle + \hat{U}\,|n\rangle - \varepsilon_n\left|N^{(1)}\right\rangle - E_n^{(1)}\,|n\rangle\right)g^1$$

$$+ .....$$

$$+ \left(\hat{H}_0\left|N^{(k)}\right\rangle + \hat{U}\left|N^{(k-1)}\right\rangle - \varepsilon_n\left|N^{(k)}\right\rangle - E_n^{(1)}\left|N^{(k-1)}\right\rangle - .... - E_n^{(k)}\,|n\rangle\right)g^k$$

$$+ .....$$

Since the power series is equal to zero, the coefficient of each power of g must be equal to zero.

We get (labelling the equation by the corresponding power of g)

$$0^{th} - order \qquad \hat{H}_0\,|n\rangle = \varepsilon_n\,|n\rangle$$

which is just our original assumption = unperturbed solution.

$$1^{st} - order \qquad \hat{H}_0\left|N^{(1)}\right\rangle + \hat{U}\,|n\rangle = \varepsilon_n\left|N^{(1)}\right\rangle + E_n^{(1)}\,|n\rangle$$

.....

.....

$$k^{th} - order \qquad \hat{H}_0\left|N^{(k)}\right\rangle + \hat{U}\left|N^{(k-1)}\right\rangle$$

$$= \varepsilon_n\left|N^{(k)}\right\rangle + E_n^{(1)}\left|N^{(k-1)}\right\rangle + ... + E_n^{(k)}\left|N^{(0)}\right\rangle$$

.....

where we have used the notation $|n\rangle = |N^{(0)}\rangle$.

Let us consider the 1^{st} – order equation.

If we apply the linear functional $\langle n|$ we get

$$\langle n|\hat{H}_0\left|N^{(1)}\right\rangle + \langle n|\hat{U}|n\rangle = \langle n|\varepsilon_n\left|N^{(1)}\right\rangle + \langle n|E_n^{(1)}|n\rangle$$

$$\varepsilon_n\left\langle n\,\middle|\,N^{(1)}\right\rangle + \langle n|\hat{U}|n\rangle = \varepsilon_n\left\langle n\,\middle|\,N^{(1)}\right\rangle + E_n^{(1)}\langle n\,|\,n\rangle$$

or since $\langle n|N^{(1)}\rangle = 0$ we get

$$E_n^{(1)} = \langle n|\hat{U}|n\rangle = 1^{st} - order \text{ correction to the energy}$$

$$= \text{ diagonal matrix element of } \hat{U} \text{ in the } |n\rangle \text{ (unperturbed) basis}$$

$$\text{or the expectation value of } \hat{U} \text{ in the state } |n\rangle$$

$$\text{or in the } n^{th} \text{ unperturbed state}$$

Therefore, to first order in g we have

$$E_n = \varepsilon_n + gE_n^{(1)} = \varepsilon_n + g\langle n|\hat{U}|n\rangle$$

$$= \varepsilon_n + \langle n|\hat{V}|n\rangle$$

where we have reabsorbed the factor g back into the original potential energy function.

In the same manner, if we apply the linear functional $\langle n|$ to the k^{th} – order equation we get

$$E_n^{(k)} = \langle n|\hat{U}\left|N^{(k-1)}\right\rangle$$

This says that, if we know the correction to the eigenvector to order $(k-1)$, then we can calculate the correction to the energy eigenvalue to order k (the next order).

Now the k^{th} –order correction to the eigenvector, $|N^{(k)}\rangle$ is just another vector in the space and, hence, we can expand it as a linear combination of the $|n\rangle$ states (since they are a basis).

$$\left|N^{(k)}\right\rangle = \sum_{m \neq n} |m\rangle \left\langle m \mid N^{(k)}\right\rangle$$

The state $|n\rangle$ is not included because $\langle n|N^{(i)}\rangle = 0$ by our choice of normalization.

In order to evaluate this sum, we must find an expression for the coefficients $\langle m|N^{(k)}\rangle$.

This can be done by applying the linear functional $\langle m|$, $m \neq n$ to the k^{th} – order equation.

We get

$$\begin{aligned}
&\langle m|\,\hat{H}_0\left|N^{(k)}\right\rangle + \langle m|\,\hat{U}\left|N^{(k-1)}\right\rangle \\
&\quad = \langle m|\,\varepsilon_n\left|N^{(k)}\right\rangle + \langle m|\,E_n^{(1)}\left|N^{(k-1)}\right\rangle + \ldots. + \langle m|\,E_n^{(k)}\left|N^{(0)}\right\rangle \\
&\varepsilon_m \left\langle m \mid N^{(k)}\right\rangle + \langle m|\,\hat{U}\left|N^{(k-1)}\right\rangle \\
&\quad = \varepsilon_n \left\langle m \mid N^{(k)}\right\rangle + E_n^{(1)} \left\langle m \mid N^{(k-1)}\right\rangle + \ldots\ldots + E_n^{(k)} \left\langle m \mid N^{(0)}\right\rangle
\end{aligned}$$

If we assume that $\varepsilon_m \neq \varepsilon_n$ (we have nondegenerate levels) we get

$$\begin{aligned}
&\left\langle m \mid N^{(k)}\right\rangle \\
&\quad = \frac{1}{\varepsilon_n - \varepsilon_m}\left(\langle m|\,\hat{U}\left|N^{(k-1)}\right\rangle - E_n^{(1)}\left\langle m \mid N^{(k-1)}\right\rangle - \ldots\ldots - E_n^{(k)}\left\langle m \mid N^{(0)}\right\rangle\right)
\end{aligned}$$

This formula allows us to find the k^{th} – order correction to the eigenvector in terms of lower order corrections to $|N\rangle$ and E_n as long as $|n\rangle$ corresponds to a nondegenerate level.

To see how this works, we will calculate the corrections to second order.

For first order, we let $k = 1$ and get

$$\begin{aligned}\left\langle m \mid N^{(1)}\right\rangle &= \frac{1}{\varepsilon_n - \varepsilon_m}\left(\langle m|\hat{U}\left|N^{(0)}\right\rangle - E_n^{(1)}\left\langle m \mid N^{(0)}\right\rangle\right) \\ &= \frac{1}{\varepsilon_n - \varepsilon_m}\left(\langle m|\hat{U}|n\rangle - E_n^{(1)}\langle m \mid n\rangle\right) \\ &= \frac{1}{\varepsilon_n - \varepsilon_m}\langle m|\hat{U}|n\rangle\end{aligned}$$

which gives

$$\left|N^{(1)}\right\rangle = \sum_{m\neq n}|m\rangle\left\langle m \mid N^{(1)}\right\rangle = \sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n - \varepsilon_m}\langle m|\hat{U}|n\rangle$$

Therefore, to first order in g we have

$$|N\rangle = |n\rangle + g\left|N^{(1)}\right\rangle = |n\rangle + \sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n - \varepsilon_m}\langle m|\hat{V}|n\rangle$$

We then use $N^{(1)}$ to calculate $E^{(2)}$, the second order correction to the energy, using

$$\begin{aligned}E_n^{(2)} &= \langle n|\hat{U}\left|N^{(1)}\right\rangle = \langle n|\hat{U}\left(\sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n - \varepsilon_m}\langle m|\hat{U}|n\rangle\right) \\ &= \sum_{m\neq n}\frac{\left|\langle n|\hat{U}|m\rangle\right|^2}{\varepsilon_n - \varepsilon_m}\end{aligned}$$

Therefore, to second order ın g we have

$$\begin{aligned}E_n &= \varepsilon_n + gE_n^{(1)} + g^2E_n^{(2)} \\ &= \varepsilon_n + \langle n|\hat{V}|n\rangle + \sum_{m\neq n}\frac{\left|\langle n|\hat{V}|m\rangle\right|^2}{\varepsilon_n - \varepsilon_m}\end{aligned}$$

We then obtain the second order correction to the state vector in the same way.

$$\begin{aligned}
\left\langle m \mid N^{(2)} \right\rangle &= \frac{1}{\varepsilon_n - \varepsilon_m} \left(\langle m| \hat{U} \left| N^{(1)} \right\rangle - E_n^{(1)} \left\langle m \mid N^{(1)} \right\rangle - E_n^{(2)} \left\langle m \mid N^{(0)} \right\rangle \right) \\
&= \frac{1}{\varepsilon_n - \varepsilon_m} \langle m| \hat{U} \left| N^{(1)} \right\rangle - \frac{1}{\varepsilon_n - \varepsilon_m} \langle n| \hat{U} |n\rangle \left\langle m \mid N^{(1)} \right\rangle \\
&= \frac{1}{\varepsilon_n - \varepsilon_m} \langle m| \hat{U} \left(\sum_{k \neq n} |k\rangle \frac{1}{\varepsilon_n - \varepsilon_k} \langle k| \hat{U} |n\rangle \right) \\
&\qquad - \frac{1}{\varepsilon_n - \varepsilon_m} \langle n| \hat{U} |n\rangle \frac{1}{\varepsilon_n - \varepsilon_m} \langle m| \hat{U} |n\rangle \\
&= \sum_{k \neq n} \frac{\langle m| \hat{U} |k\rangle \langle k| \hat{U} |n\rangle}{(\varepsilon_n - \varepsilon_m)(\varepsilon_n - \varepsilon_k)} - \frac{\langle n| \hat{U} |n\rangle \langle m| \hat{U} |n\rangle}{(\varepsilon_n - \varepsilon_m)^2}
\end{aligned}$$

Therefore,

$$\left| N^{(2)} \right\rangle = \sum_{m \neq n} \sum_{k \neq n} |m\rangle \frac{\langle m| \hat{U} |k\rangle \langle k| \hat{U} |n\rangle}{(\varepsilon_n - \varepsilon_m)(\varepsilon_n - \varepsilon_k)} - \sum_{m \neq n} |m\rangle \frac{\langle n| \hat{U} |n\rangle \langle m| \hat{U} |n\rangle}{(\varepsilon_n - \varepsilon_m)^2}$$

and so on.

An Example

We will now do an example where we know the exact answer so that we can compare it to the perturbation results.

We consider a 1-dimensional system represented by a perturbed harmonic oscillator where

$$\hat{H} = \hat{H}_0 + \hat{V}$$

with

$$\hat{H}_0 = \hbar\omega(\hat{a}^+\hat{a} + \frac{1}{2}) \rightarrow \text{ harmonic oscillator}$$

$$\hat{H}_0 |n\rangle = \varepsilon_n |n\rangle = \hbar\omega(n + \frac{1}{2}) |n\rangle$$

In standard operator notation

$$\hat{H}_0 = \frac{\hat{p}^2}{2m} + \frac{1}{2}kx^2 \quad , \quad k = m\omega^2$$

We now perturb the system with the potential energy term

$$\hat{V} = \frac{1}{2}k'x^2 \quad , \quad k' << k$$

Therefore

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}(k + k')x^2$$

We still have a harmonic oscillator (with a changed spring constant).

This says that the new energies are given by

$$E_n = \hbar\tilde{\omega}(n + \frac{1}{2})$$

where

$$\tilde{\omega} = \sqrt{\frac{k + k'}{m}} = \sqrt{\frac{k}{m}}\sqrt{1 + \frac{k'}{k}} = \omega\sqrt{1 + \frac{k'}{k}}$$

Therefore, the exact energies for the perturbed system are

$$E_n = \hbar\tilde{\omega}(n + \frac{1}{2}) = \hbar\omega\sqrt{1 + \frac{k'}{k}}(n + \frac{1}{2})$$

For $k' \ll k$, we can expand this as

$$E_n = \hbar\omega(n+\frac{1}{2})\left(1+\frac{1}{2}\frac{k'}{k}-\frac{1}{8}\left(\frac{k'}{k}\right)^2+\ldots\right)$$

which should correspond to the perturbation calculated energy calculated to 2nd order in perturbation theory.

We now do the perturbation calculation.

Our earlier derivation gives

$$E_n = \varepsilon_n + \langle n|\hat{V}|n\rangle + \sum_{m\neq n}\frac{\left|\langle n|\hat{V}|m\rangle\right|^2}{\varepsilon_n-\varepsilon_m}$$

$$|N\rangle = |n\rangle + \sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n-\varepsilon_m}\langle m|\hat{V}|n\rangle$$

where

$$\hat{V} = \frac{1}{2}k'x^2 = \frac{1}{4}\frac{k'\hbar}{m\omega}(\hat{a}+\hat{a}^+)^2$$

We need to calculate this matrix element

$$\begin{aligned}\langle m|\hat{V}|n\rangle &= \frac{1}{4}\frac{k'\hbar}{m\omega}\langle m|(\hat{a}+\hat{a}^+)^2|n\rangle = \frac{1}{4}\frac{k'\hbar}{m\omega}\langle m|\hat{a}^2+\hat{a}\hat{a}^+ +\hat{a}^+\hat{a}+(\hat{a}^+)^2|n\rangle \\ &= \frac{1}{4}\frac{k'\hbar}{m\omega}\langle m|\left(\sqrt{n(n-1)}|n-2\rangle+(n+1)|n\rangle+n|n\rangle+\sqrt{(n+1)(n+2)}|n+2\rangle\right) \\ &= \frac{1}{4}\frac{k'\hbar}{m\omega}\left(\sqrt{n(n-1)}\delta_{m,n-2}+(2n+1)\delta_{m,n}+\sqrt{(n+1)(n+2)}\delta_{m,n+2}\right)\end{aligned}$$

where we have used

$$\langle m|\,\hat{a}\,|n\rangle = \sqrt{n}\,\langle m\mid n-1\rangle = \sqrt{n}\delta_{m,n-1}$$

$$\langle m|\,\hat{a}^{+}\,|n\rangle = \sqrt{n+1}\,\langle m\mid n+1\rangle = \sqrt{n+1}\delta_{m,n+1}$$

Therefore

$$\langle n|\,\hat{V}\,|n\rangle = \frac{1}{4}\frac{k'\hbar}{m\omega}(2n+1) = \hbar\omega(n+\frac{1}{2})\left(\frac{1}{2}\frac{k'}{k}\right)$$

and

$$\sum_{m\neq n}\frac{\left|\langle n|\,\hat{V}\,|m\rangle\right|^2}{\varepsilon_n-\varepsilon_m} = \left(\frac{1}{4}\frac{k'\hbar}{m\omega}\right)^2\left[\frac{n(n-1)}{2\hbar\omega}-\frac{(n+1)(n+2)}{(-2\hbar\omega)}\right]$$

$$= -\hbar\omega\left(n+\frac{1}{2}\right)\left[\frac{1}{8}\left(\frac{k'}{k}\right)^2\right]$$

which then gives

$$E_n = \hbar\omega(n+\frac{1}{2})\left(1+\frac{1}{2}\frac{k'}{k}-\frac{1}{8}\left(\frac{k'}{k}\right)^2+\ldots\right)$$

in agreement with the exact result (to 2nd order).

To calculate the new state vector to first order we need

$$\sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n-\varepsilon_m}\langle m|\,\hat{V}\,|n\rangle$$

$$= \frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{n(n-1)}}{2\hbar\omega}|n-2\rangle + \frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{(n+1)(n+2)}}{(-2\hbar\omega)}|n+2\rangle$$

which gives

$$|N\rangle = |n\rangle + \frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{n(n-1)}}{2\hbar\omega}|n-2\rangle - \frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{(n+1)(n+2)}}{2\hbar\omega}|n+2\rangle$$

What does the new ground state wave function look like?

We have

$$|N=0\rangle = |0\rangle - \frac{\sqrt{2}}{2\hbar\omega}|2\rangle$$

and

$$\langle x \mid N=0\rangle = \langle x \mid 0\rangle - \frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{2}}{2\hbar\omega}\langle x \mid 2\rangle$$

$$\psi_{N=0}(x) = \psi_0(x) - \frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{2}}{2\hbar\omega}\psi_2(x)$$

Now we found earlier that

$$\langle x \mid 0\rangle = \psi_0(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega x^2}{2\hbar}}$$

and

$$\langle x \mid 2\rangle = \psi_2(x) = \left(\frac{m\omega}{4\pi\hbar}\right)^{1/4} (2\frac{m\omega}{\hbar}x^2 - 1)e^{-\frac{m\omega x^2}{2\hbar}}$$

which gives

$$\psi_{N=0}(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega x^2}{2\hbar}}\left(1 + \frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{2}}{2\hbar\omega}\frac{1}{\sqrt{2}}\left(1 - 2\frac{m\omega}{\hbar}x^2\right)\right)$$

$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega x^2}{2\hbar}}\left(1 + \frac{k'}{8k} - \frac{m\omega k'}{4\hbar k}x^2\right)$$

Since we are only changing the spring constant we should have

$$\psi_{N=0}(x) = \left(\frac{m\tilde{\omega}}{\pi\hbar}\right)^{1/4} e^{-\frac{m\tilde{\omega}x^2}{2\hbar}} = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \left(1+\frac{k'}{k}\right)^{1/8} e^{-\frac{m\omega x^2}{2\hbar}\sqrt{1+\frac{k'}{k}}}$$

$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \left(1+\frac{k'}{8k}\right) e^{-\frac{m\omega x^2}{2\hbar}\left(1+\frac{k'}{2k}\right)}$$

$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \left(1+\frac{k'}{8k}\right)\left(1-\frac{m\omega k'}{4\hbar k}x^2\right) e^{-\frac{m\omega x^2}{2\hbar}}$$

$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega x^2}{2\hbar}} \left(1+\frac{k'}{8k}-\frac{m\omega k'}{4\hbar k}x^2\right)$$

which agrees with the perturbation result to this order.

The perturbation theory we have developed so far breaks down if there are any states where

$$\varepsilon_n = \varepsilon_m \text{ but } \langle m|V|n\rangle \neq 0$$

i.e., degenerate states with nonzero matrix elements of the perturbing potential between them.

Degenerate Case

We handle this case as follows.

Suppose we have a group of *k* states

$$|n_1\rangle \ , \ |n_2\rangle \ , \ |n_3\rangle \ , \ \ldots \ , \ |n_k\rangle$$

that are degenerate states of the unperturbed Hamiltonian $\hat{H}_0$, i.e.,

$$\hat{H}_0|n_i\rangle = \varepsilon_{n_1}|n_i\rangle \ , \ i = 1,2,3,4,5,\ldots,k$$

If $\langle n_i|\hat{V}|n_j\rangle \neq 0$ for $i \neq j$ within this set, the previous perturbation formulas will fail because the energy denominators $\varepsilon_{ni} - \varepsilon_{nj} \to 0$.

Remember, however, that any linear combination of the degenerate states

$$|n_1\rangle \ , \ |n_2\rangle \ , \ |n_3\rangle \ , \ \ldots \ , \ |n_k\rangle$$

is also an eigenstate of $\hat{H}_0$ with the same energy ε_{n1} .

Therefore, if we can choose a different set of basis states (start off with a new set of *zero-order states*) within this degenerate subspace, i.e., choose a new set of k orthogonal states (linear combinations of the old set of degenerate states)

$$|n_\alpha\rangle = \sum_{i=1}^{k} C_{\alpha i} |n_i\rangle$$

such that we have

$$\langle n_\alpha|\hat{V}|n_\beta\rangle = 0 \text{ for } \alpha \neq \beta$$

then we can use the perturbation formulas as derived earlier.

This procedure will work because the terms with zero denominators will have zero numerators and if one looks at the derivation, this means that these terms do not even appear in the final results, i.e., the zero numerators take effect before the zero denominators appear.

This condition says that the correct choice of zero-order states *within the degenerate subspace* (*the set of degenerate vectors*) for doing degenerate perturbation theory is that set which diagonalizes the matrix representation of $\hat{V}$ within each group of degenerate states.

The problem of diagonalizing $\hat{V}$ within a group of k states

$$|n_1\rangle \ , \ |n_2\rangle \ , \ |n_3\rangle \ , \ \ldots \ , \ |n_k\rangle$$

is that of finding the eigenvectors and eigenvalues of the $k \times k$ matrix

$$\begin{pmatrix} \langle n_1|\hat{V}|n_1\rangle & \langle n_1|\hat{V}|n_2\rangle & \cdot & \langle n_1|\hat{V}|n_k\rangle \\ \langle n_2|\hat{V}|n_1\rangle & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \langle n_k|\hat{V}|n_1\rangle & \cdot & \cdot & \langle n_k|\hat{V}|n_k\rangle \end{pmatrix}$$

We now show that if the coefficients $C_{\alpha i}$ of the new zero-order states are just the components of the eigenvectors of this matrix, then it will be diagonalized in the degenerate subspace.

Suppose that we represent the eigenvector by the column vector

$$|n_\alpha\rangle = \begin{pmatrix} C_{\alpha 1} \\ C_{\alpha 2} \\ \cdot \\ C_{\alpha k} \end{pmatrix}$$

then the statement that $|n_\alpha\rangle$ is an eigenvector of the $k \times k$ submatrix of $\hat{V}$ with eigenvalues that we write as $E^{(1)}$ is equivalent to writing

$$\hat{V}|n_\alpha\rangle = E^{(1)}_{n_\alpha}|n_\alpha\rangle$$

or

$$\begin{pmatrix} \langle n_1|\hat{V}|n_1\rangle & \langle n_1|\hat{V}|n_2\rangle & \cdot & \langle n_1|\hat{V}|n_k\rangle \\ \langle n_2|\hat{V}|n_1\rangle & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \langle n_k|\hat{V}|n_1\rangle & \cdot & \cdot & \langle n_k|\hat{V}|n_k\rangle \end{pmatrix} \begin{pmatrix} C_{\alpha 1} \\ C_{\alpha 2} \\ \cdot \\ C_{\alpha k} \end{pmatrix} = E^{(1)}_{n_\alpha} \begin{pmatrix} C_{\alpha 1} \\ C_{\alpha 2} \\ \cdot \\ C_{\alpha k} \end{pmatrix}$$

or finally

$$\sum_i \langle n_j|\hat{V}|n_i\rangle C_{\alpha i} = E^{(1)}_{n_\alpha} C_{\alpha i}$$

All of these calculations take place within the k-dimensional degenerate subspace.

We will assume that the eigenvectors are normalized, which implies that

$$\sum_i |C_{\alpha i}|^2 = 1$$

i.e., vectors are normalized to one.

Now consider another of the new vectors given by

$$|n_\beta\rangle = \sum_{j=1}^{k} C_{\beta j} |n_j\rangle$$

We then have

$$\langle n_\beta| = \sum_{j=1}^{k} C^*_{\beta j} \langle n_j|$$

Applying the linear functional $\langle n_\beta|$ to the eigenvector/eigenvalue equation we get

$$\sum_j \sum_i \langle n_j | C^*_{\beta j} \hat{V} C_{\alpha i} | n_i \rangle = E^{(1)}_{n_\alpha} \sum_j C^*_{\beta j} C_{\alpha i}$$

Now, since the eigenvectors of any Hermitian matrix are always a complete orthonormal set (or can always be made so using the Gram-Schmidt process), the orthonormality of the new vectors says that

$$\sum_j C^*_{\beta j} C_{\alpha i} = \langle n_\beta \mid n_\alpha \rangle = \delta_{\beta\alpha}$$

Therefore the vectors

$$|n_\alpha\rangle = \sum_{i=1}^{k} C_{\alpha i} |n_i\rangle$$

satisfy

$$\left(\sum_j \langle n_j | C^*_{\beta j} \right) \hat{V} \left(\sum_i C_{\alpha i} |n_i\rangle \right) = E^{(1)}_{n_\alpha} \delta_{\alpha\beta}$$

$$\langle n_\beta | \hat{V} | n_\alpha \rangle = E^{(1)}_{n_\alpha} \delta_{\alpha\beta}$$

This says that the corresponding eigenvalue $E^{(1)}{}_{n\alpha}$ of one of the new vectors is the first order energy corrections for the state $|N_\alpha\rangle$.

The states

$$|n_\alpha\rangle \ , \ |n_\beta\rangle \ , \ |n_\gamma\rangle \ , \ \ldots \ , \ |n_\kappa\rangle$$

are called the *new zeroth-order state vectors*.

Thus the group of states

$$|n_1\rangle \ , \ |n_2\rangle \ , \ |n_3\rangle \ , \ \ldots \ , \ |n_k\rangle$$

in the presence of the perturbation $\hat{V}$ split(rearrange) into the *k* states

$$|n_\alpha\rangle \ , \ |n_\beta\rangle \ , \ |n_\gamma\rangle \ , \ \ldots \ , \ |n_\kappa\rangle$$

which are given to first order by

$$|N_\alpha\rangle = |n_\alpha\rangle + \sum_{m \neq \alpha,\beta,..,\kappa} \frac{|m\rangle \langle m| \hat{V} |n_\alpha\rangle}{\varepsilon_{n_1} - \varepsilon_m}$$

and the energy shift to second order is

$$E_{n_\alpha} = \varepsilon_{n_1} + \langle n_\alpha| \hat{V} |n_\alpha\rangle + \sum_{m \neq \alpha,\beta,..,\kappa} \frac{\left|\langle m| \hat{V} |n_\alpha\rangle\right|^2}{\varepsilon_{n_1} - \varepsilon_m}$$

where

$$\langle n_\alpha| \hat{V} |n_\alpha\rangle = E^{(1)}_{n_\alpha}$$

is an eigenvalue of the $\hat{V}$ matrix in the degenerate subspace.

An Example

We now consider a 2–dimensional oscillator that is perturbed by a potential of the form

$$\hat{V} = \lambda \hat{x} \hat{y}$$

We then have

$$\hat{H} = \hat{H}_0 + \hat{V}$$

where

$$\hat{H}_0 = \frac{\hat{p}_x^2}{2m} + \frac{\hat{p}_y^2}{2m} + \frac{1}{2} k (x^2 + y^2)$$

As we showed earlier, using the $\hat{a}_x$ and $\hat{a}_y$ operators we get

$$\hat{H}_0 = \hbar\omega(\hat{a}_x^+\hat{a}_x + \hat{a}_y^+\hat{a}_y + 1)$$

$$\hat{H}_0 |n_x, n_y\rangle = \varepsilon_{n_x,n_y} |n_x, n_y\rangle = \hbar\omega(n_x + n_y + 1) |n_x, n_y\rangle$$

$$\text{degeneracy} = n_x + n_y + 1$$

and

$$\hat{V} = \lambda\frac{\hbar}{2m\omega}(\hat{a}_x + \hat{a}_x^+)(\hat{a}_y + \hat{a}_y^+)$$

The unperturbed ground-state is $|0,0\rangle$ with $\varepsilon_{0,0} = \hbar\omega$.

It is a nondegenerate level, so we can apply the standard perturbation theory to get

$$E_0 = \varepsilon_{0,0} + \langle 0,0|\hat{V}|0,0\rangle + \sum_{m\neq 0}\frac{|\langle m,n|\hat{V}|0,0\rangle|}{\varepsilon_{0,0} - \varepsilon_{m,n}}$$

Now

$$\langle 0,0|\hat{V}|0,0\rangle = \frac{\lambda\hbar}{2m\omega}\langle 0,0|(\hat{a}_x + \hat{a}_x^+)(\hat{a}_y + \hat{a}_y^+)|0,0\rangle = 0$$

and

$$\langle m,n|\hat{V}|0,0\rangle = \frac{\lambda\hbar}{2m\omega}\langle m,n|(\hat{a}_x + \hat{a}_x^+)(\hat{a}_y + \hat{a}_y^+)|0,0\rangle$$

$$= \frac{\lambda\hbar}{2m\omega}\langle m,n|\hat{a}_x^+\hat{a}_y^+|0,0\rangle = \frac{\lambda\hbar}{2m\omega}\langle m,n \mid 1,1\rangle$$

$$= \frac{\lambda\hbar}{2m\omega}\delta_{m,1}\delta_{n,1}$$

Thus, the correction to first order is zero.

Calculating to second order we get

$$E_0 = \hbar\omega - \left(\frac{\lambda\hbar}{2m\omega}\right)^2 \frac{1}{2\hbar\omega} = \hbar\omega\left(1 - \frac{\lambda^2}{8m^2\omega^4}\right)$$

The next unperturbed level is 2–fold degenerate, i.e.,

$$n_x = 0, n_y = 1 \rightarrow \varepsilon_{0,1} = 2\hbar\omega$$

$$n_x = 1, n_y = 0 \rightarrow \varepsilon_{1,0} = 2\hbar\omega$$

For ease of notation we will sometimes denote

$$|1,0\rangle = |a\rangle \text{ and } |0,1\rangle = |b\rangle$$

We now use degenerate perturbation theory.

The procedure is to evaluate the $\hat{V}$ matrix in the 2×2 degenerate subspace, diagonalize it and obtain the first order corrections.

The 2×2 matrix is

$$V = \begin{pmatrix} \hat{V}_{aa} & \hat{V}_{ab} \\ \hat{V}_{ba} & \hat{V}_{bb} \end{pmatrix} = \begin{pmatrix} \langle a|\hat{V}|a\rangle & \langle a|\hat{V}|b\rangle \\ \langle b|\hat{V}|a\rangle & \langle b|\hat{V}|b\rangle \end{pmatrix}$$

Now

$$\hat{V}_{aa} = \langle 1,0|\hat{V}|1,0\rangle = 0 = \langle 0,1|\hat{V}|0,1\rangle = \hat{V}_{bb}$$

$$\hat{V}_{ab} = \hat{V}_{ba} = \langle 1,0|\hat{V}|0,1\rangle = \frac{\lambda\hbar}{2m\omega}\langle 1,0|\hat{a}_x^+\hat{a}_y|0,1\rangle$$

$$= \frac{\lambda\hbar}{2m\omega}$$

Therefore the 2×2 submatrix is

$$V = \frac{\lambda\hbar}{2m\omega}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

This is simple to diagonalize.

We get these results

$$|a'\rangle = \frac{1}{\sqrt{2}}(|a\rangle + |b\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow \text{ eigenvalue } = +\frac{\lambda\hbar}{2m\omega}$$

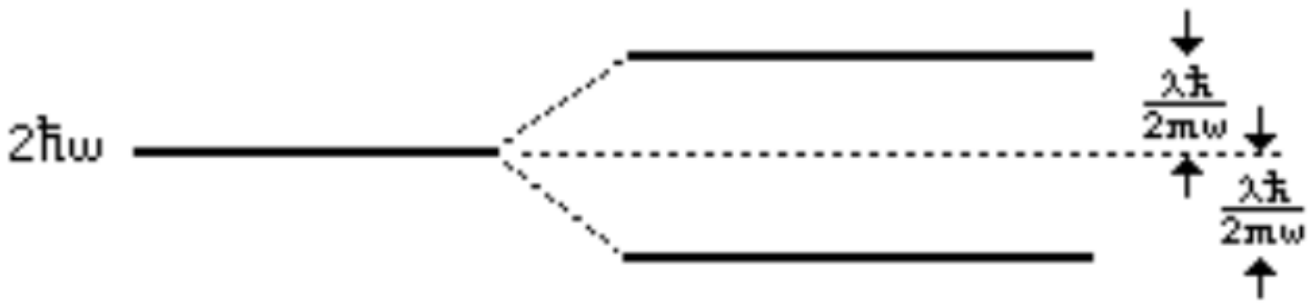
$$|b'\rangle = \frac{1}{\sqrt{2}}(|a\rangle - |b\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \rightarrow \text{ eigenvalue } = -\frac{\lambda\hbar}{2m\omega}$$

$|a'\rangle$ and $|b'\rangle$ are the new zeroth order state vectors (eigenvectors of the 2×2 submatrix) and

$$\pm\frac{\lambda\hbar}{2m\omega}$$

are the corresponding first order energy corrections.

Thus, the 2–fold degenerate level splits into two nondegenerate levels as shown in the Figure below.



Splitting of a degenerate Level

where

$$E_{a'} = 2\hbar\omega - \frac{\lambda\hbar}{2m\omega}$$

$$E_{b'} = 2\hbar\omega + \frac{\lambda\hbar}{2m\omega}$$

$$\Delta E = \text{levelsplitting} = \frac{\lambda\hbar}{m\omega}$$

Another Example

Now let us consider a system of two spin−1/2 particles in a magnetic field.

We also assume that there exists a direct spin-spin interaction so that the Hamiltonian takes the form

$$\hat{H} = \left(\alpha \vec{S}_{1,op} + \beta \vec{S}_{2,op}\right) \cdot \vec{B} + \gamma \vec{S}_{1,op} \cdot \vec{S}_{2,op}$$

If we choose $\vec{\mathrm{B}} = B\hat{z}$ and $\gamma \gg \alpha, \beta$, then we can write

$$\hat{H} = \hat{H}_0 + \hat{V}$$

$$\hat{H}_0 = \gamma \vec{S}_{1,op} \cdot \vec{S}_{2,op}$$

$$\hat{V} = \alpha B \hat{S}_{1z} + \beta B \hat{S}_{2z}$$

We define

$$\vec{S}_{op} = \vec{S}_{1,op} + \vec{S}_{2,op} = \text{ total spin angular momentum}$$

and then we have

$$\begin{aligned} \vec{S}_{op}^2 = \vec{S}_{op} \cdot \vec{S}_{op} &= (\vec{S}_{1,op} + \vec{S}_{2,op}) \cdot (\vec{S}_{1,op} + \vec{S}_{2,op}) \\ &= \vec{S}_{1,op}^2 + \vec{S}_{2,op}^2 + 2\vec{S}_{1,op} \cdot \vec{S}_{2,op} \\ &= \frac{3}{4}\hbar^2 \hat{I} + \frac{3}{4}\hbar^2 \hat{I} + 2\vec{S}_{1,op} \cdot \vec{S}_{2,op} \end{aligned}$$

or

$$\hat{H}_0 = \gamma \vec{S}_{1,op} \cdot \vec{S}_{2,op} = \frac{\gamma}{2}\left(\vec{S}_{op}^2 - \frac{3}{2}\hbar^2 \hat{I}\right)$$

Our earlier discussion of the addition of angular momentum says that when we add two spin−1/2 angular momenta we get the resultant total angular momentum values 0 and 1, i.e.,

$$\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$$

Each separate spin−1/2 system has the eigenvectors/eigenvalues

$$\vec{S}^2_{1,op} |\pm\rangle = \frac{3}{4}\hbar |\pm\rangle \quad , \quad \vec{S}_{1z} |\pm\rangle = \pm\frac{\hbar}{2} |\pm\rangle$$

The corresponding direct-product states are

$$|{+}{+}\rangle \ , \ |{+}{-}\rangle \ , \ |{-}{+}\rangle \ , \ |{-}{-}\rangle$$

where the symbols mean

$$|{+}{-}\rangle = |+\rangle_1 |-\rangle_2$$

and so on.

The total angular momentum states are (we derived them earlier) labeled as $|s,m\rangle$ where

$$\vec{S}^2_{op} |s,m\rangle = \hbar^2 s(s+1) |s,m\rangle$$
$$\hat{S}_z |s,m\rangle = (\hat{S}_{1z} + \hat{S}_{2z}) |s,m\rangle = \pm m\hbar |s,m\rangle$$

They are given in terms of the direct product states by

$$|1,1\rangle = |{+}{+}\rangle = |1\rangle$$
$$|1,0\rangle = \frac{1}{\sqrt{2}} |{+}{-}\rangle + \frac{1}{\sqrt{2}} |{-}{+}\rangle = |2\rangle$$
$$|1,-1\rangle = |{-}{-}\rangle = |3\rangle$$
$$|0,0\rangle = \frac{1}{\sqrt{2}} |{+}{-}\rangle - \frac{1}{\sqrt{2}} |{-}{+}\rangle = |4\rangle$$

The total angular momentum states are eigenstates of $\hat{H}_0$ and we use them as the unperturbed or zero-order states.

$$\hat{H}_0 |1,1\rangle = \frac{\gamma\hbar^2}{4} |1,1\rangle = \varepsilon_1 |1,1\rangle$$

$$\hat{H}_0 |1,0\rangle = \frac{\gamma\hbar^2}{4} |1,0\rangle = \varepsilon_2 |1,0\rangle$$

$$\hat{H}_0 |1,-1\rangle = \frac{\gamma\hbar^2}{4} |1,-1\rangle = \varepsilon_3 |1,-1\rangle$$

$$\hat{H}_0 |0,0\rangle = -\frac{3\gamma\hbar^2}{4} |0,0\rangle = \varepsilon_4 |0,0\rangle$$

We thus have one nondegenerate level and one 3−fold degenerate level.

Now using

$$\hat{V} = \alpha B \hat{S}_{1z} + \beta B \hat{S}_{2z}$$

we do perturbation theory on these levels.

Nondegenerate Level

First order:

$$E_4^{(1)} = \langle 4| \hat{V} |4\rangle$$

$$= \left(\frac{1}{\sqrt{2}} \langle +-| + \frac{1}{\sqrt{2}} \langle -+| \right) (\alpha B \hat{S}_{1z} + \beta B \hat{S}_{2z}) \left(\frac{1}{\sqrt{2}} |+-\rangle - \frac{1}{\sqrt{2}} |-+\rangle \right) = 0$$

Second order:

$$E_4^{(2)} = \sum_{m\neq 4} \frac{\left|\langle m|\hat{V}|4\rangle\right|^2}{\varepsilon_4-\varepsilon_m} = \frac{\left|\langle ++|\,\alpha B\hat{S}_{1z}+\beta B\hat{S}_{2z}\left(\frac{1}{\sqrt{2}}|+-\rangle-\frac{1}{\sqrt{2}}|-+\rangle\right)\right|^2}{\varepsilon_4-\varepsilon_1}$$
$$+\frac{\left|\left(\frac{1}{\sqrt{2}}\langle +-|+\frac{1}{\sqrt{2}}\langle -+|\right)\alpha B\hat{S}_{1z}+\beta B\hat{S}_{2z}\left(\frac{1}{\sqrt{2}}|+-\rangle-\frac{1}{\sqrt{2}}|-+\rangle\right)\right|^2}{\varepsilon_4-\varepsilon_2}$$
$$+\frac{\left|\langle --|\,\alpha B\hat{S}_{1z}+\beta B\hat{S}_{2z}\left(\frac{1}{\sqrt{2}}|+-\rangle-\frac{1}{\sqrt{2}}|-+\rangle\right)\right|^2}{\varepsilon_4-\varepsilon_3}$$

$$E_4^{(2)} = \frac{\left|\left(\frac{1}{\sqrt{2}}\langle +-|+\frac{1}{\sqrt{2}}\langle -+|\right)\alpha B\hat{S}_{1z}+\beta B\hat{S}_{2z}\left(\frac{1}{\sqrt{2}}|+-\rangle-\frac{1}{\sqrt{2}}|-+\rangle\right)\right|^2}{\varepsilon_4-\varepsilon_2}$$
$$= -\frac{B^2(\alpha-\beta)^2}{\gamma}$$

Therefore the energy to second order for the non-degenerate level is

$$E_4 = \varepsilon_4 + E_4^{(1)} + E_4^{(2)} = -\frac{3\gamma\hbar^2}{4} - \frac{B^2(\alpha-\beta)^2}{\gamma}$$

Degenerate Level

In this case, the 3×3 degenerate submatrix of $\hat{V}$ is

$$\begin{pmatrix} \langle 1|\hat{V}|1\rangle & \langle 1|\hat{V}|2\rangle & \langle 1|\hat{V}|3\rangle \\ \langle 2|\hat{V}|1\rangle & \langle 2|\hat{V}|2\rangle & \langle 2|\hat{V}|3\rangle \\ \langle 3|\hat{V}|1\rangle & \langle 3|\hat{V}|2\rangle & \langle 3|\hat{V}|3\rangle \end{pmatrix} = \frac{(\alpha+\beta)\hbar B}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

which is already diagonal.

Since the diagonal elements are the first order energy corrections, we have (to first order)

$$E_1 = \frac{\gamma\hbar^2}{4} + \frac{(\alpha+\beta)\hbar B}{2}$$

$$E_2 = \frac{\gamma\hbar^2}{4}$$

$$E_3 = \frac{\gamma\hbar^2}{4} - \frac{(\alpha+\beta)\hbar B}{2}$$

Exact Solution

We can, in fact, solve this problem exactly and compare it to the perturbation result.

We do this by choosing a new basis set (arbitrary choice made to simplify calculations) and rewriting $\hat{H}$ in terms of operators appropriate to the basis choice(that is what is meant by simplify calculations).

We then use the new basis to construct the 4×4 $\hat{H}$ matrix and then diagonalize the matrix.

This method always works for a system with a small number of states.

Choose the direct product states as a basis

$$|{+}{+}\rangle = |1\rangle \ , \ |{+}{-}\rangle = |2\rangle \ , \ |{-}{+}\rangle = |3\rangle \ , \ |{-}{-}\rangle = |4\rangle$$

Write $\hat{H}$ as (choose operators appropriate(easy to calculate) to the basis or the HOME space)

$$\hat{H} = \alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z} + \gamma\left(\hat{S}_{1z}\hat{S}_{2z} + \hat{S}_{1x}\hat{S}_{2x} + \hat{S}_{1y}\hat{S}_{2y}\right)$$

$$= \alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z} + \gamma\left(\hat{S}_{1z}\hat{S}_{2z} + \frac{1}{2}\left(\hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+}\right)\right)$$

Construct the 4×4 $\hat{H}$ matrix

$$\begin{pmatrix} \langle 1|\hat{H}|1\rangle & \langle 1|\hat{H}|2\rangle & \langle 1|\hat{H}|3\rangle & \langle 1|\hat{H}|4\rangle \\ \langle 2|\hat{H}|1\rangle & \langle 2|\hat{H}|2\rangle & \langle 2|\hat{H}|3\rangle & \langle 2|\hat{H}|4\rangle \\ \langle 3|\hat{H}|1\rangle & \langle 3|\hat{H}|2\rangle & \langle 3|\hat{H}|3\rangle & \langle 3|\hat{H}|4\rangle \\ \langle 4|\hat{H}|1\rangle & \langle 4|\hat{H}|2\rangle & \langle 4|\hat{H}|3\rangle & \langle 4|\hat{H}|4\rangle \end{pmatrix}$$

Using

$$\hat{S}_z|\pm\rangle = \pm\frac{\hbar}{2}|\pm\rangle$$

$$\hat{S}_+|+\rangle = 0 = \hat{S}_-|-\rangle$$

$$\hat{S}_+|-\rangle = \hbar|+\rangle \text{ and } \hat{S}_-|+\rangle = \hbar|-\rangle$$

we get

$$\begin{pmatrix} B\hbar(\alpha+\beta)+\frac{\gamma\hbar^2}{4} & 0 & 0 & 0 \\ 0 & B\hbar(\alpha-\beta)-\frac{\gamma\hbar^2}{4} & \frac{\gamma\hbar^2}{2} & 0 \\ 0 & \frac{\gamma\hbar^2}{2} & -B\hbar(\alpha-\beta)-\frac{\gamma\hbar^2}{4} & 0 \\ 0 & 0 & 0 & -B\hbar(\alpha+\beta)+\frac{\gamma\hbar^2}{4} \end{pmatrix}$$

Diagonalizing to get the eigenvalues we find the exact energies

$$E_1 = \frac{\gamma\hbar^2}{4} + B\hbar(\alpha+\beta) \quad , \quad E_2 = -\frac{\gamma\hbar^2}{4} + \frac{1}{2}\sqrt{\gamma^2\hbar^4 + 4B^2\hbar^2(\alpha-\beta)^2}$$

$$E_3 = -\frac{\gamma\hbar^2}{4} - \frac{1}{2}\sqrt{\gamma^2\hbar^4 + 4B^2\hbar^2(\alpha-\beta)^2} \quad , \quad E_4 = \frac{\gamma\hbar^2}{4} - B\hbar(\alpha+\beta)$$

To compare to the perturbation calculation we let $B \to 0$ and we get the approximation

$$E_1 = \frac{\gamma\hbar^2}{4} + B\hbar(\alpha+\beta) \quad , \quad E_2 = \frac{\gamma\hbar^2}{4} + \frac{B^2(\alpha-\beta)^2}{\gamma}$$

$$E_3 = -\frac{3\gamma\hbar^2}{4} - \frac{B^2(\alpha-\beta)^2}{\gamma} \quad , \quad E_4 = \frac{\gamma\hbar^2}{4} - B\hbar(\alpha+\beta)$$

which agrees with the perturbation results.

More Ideas about Perturbation Methods

The main problem with Rayleigh-Schrodinger perturbation theory (RSPT) is that the form of the higher order terms becomes increasingly complex and, hence, the series is difficult to evaluate.

The Brillouin-Wigner Method

This technique allows us to see the higher order structure of the perturbation series more clearly.

Consider the energy eigenvalue equation

$$\hat{H}\,|N\rangle = E_n\,|N\rangle = \left(\hat{H}_0 + g\hat{U}\right)|N\rangle$$

Applying the linear functional $\langle m|$ we get

$$\langle m|\,\hat{H}\,|N\rangle = E_n\,\langle m \mid N\rangle = \langle m|\,\hat{H}_0\,|N\rangle + g\,\langle m|\,\hat{U}\,|N\rangle$$

$$E_n\,\langle m \mid N\rangle = \varepsilon_m\,\langle m \mid N\rangle + g\,\langle m|\,\hat{U}\,|N\rangle$$

$$(E_n - \varepsilon_m)\,\langle m \mid N\rangle = g\,\langle m|\,\hat{U}\,|N\rangle$$

We will use the normalization $\langle n \mid N\rangle = 1$ once again.

Now since the $|m\rangle$ states are a complete orthonormal basis we can always write

$$|N\rangle = \sum_m |m\rangle \langle m \mid N\rangle = |n\rangle \langle n \mid N\rangle + \sum_{m\neq n} |m\rangle \langle m \mid N\rangle$$
$$= |n\rangle + \sum_{m\neq n} |m\rangle \langle m \mid N\rangle$$

Using these results we get

$$|N\rangle = |n\rangle + \sum_{m\neq n} |m\rangle \frac{g}{E_n - \varepsilon_m} \langle m| \hat{U} |N\rangle$$

We now develop a series expansion of $|N\rangle$ in powers of g as follows:

$0^{th} - order:$

$$|N\rangle = |n\rangle$$

$1^{st} - order:$ (substitute $0^{th} - order$ result for $|N\rangle$ into general formula)

$$|N\rangle = |n\rangle + \sum_{m\neq n} |m\rangle \frac{g}{E_n - \varepsilon_m} \langle m| \hat{U} |n\rangle$$

This is not the same result as in RSPT since the full energy E_n remains in the denominator.

$2^{nd} - order:$ (substitute $1^{st} - order$ result for $|N\rangle$ into general formula)

$$|N\rangle = |n\rangle + \sum_{m\neq n} |m\rangle \frac{g}{E_n - \varepsilon_m} \langle m| \hat{U} \left(|n\rangle + \sum_{j\neq n} |j\rangle \frac{g}{E_n - \varepsilon_j} \langle j| \hat{U} |n\rangle \right)$$
$$= |n\rangle + \sum_{m\neq n} |m\rangle \frac{g}{E_n - \varepsilon_m} \langle m| \hat{U} |n\rangle$$
$$+ g^2 \sum_{m\neq n} \sum_{j\neq n} |m\rangle \frac{g}{E_n - \varepsilon_m} \langle m| \hat{U} |j\rangle \frac{g}{E_n - \varepsilon_j} \langle j| \hat{U} |n\rangle$$

and so on.

This is a complex power series in g since the full energy E_n remains in the denominator.

If we let $|m\rangle = |n\rangle$ in the general state expression we get

$$E_n \langle n \mid N \rangle = \varepsilon_n \langle n \mid N \rangle + g \langle n | \hat{U} | N \rangle$$

$$E_n = \varepsilon_n + g \langle n | \hat{U} | N \rangle$$

This can be expanded to give E_n as a series in powers of g, i.e., substituting the 0^{th} – order approximation for $|N\rangle$ gives the 1^{st} – order approximation for E_n and substituting the 1^{st} – order approximation for $|N\rangle$ gives the 2^{nd} – order approximation for E_n and so on.

If we substitute the 1^{st} – order approximation for $|N\rangle$ we get

$$E_n = \varepsilon_n + g \langle n | \hat{U} | N \rangle$$

$$= \varepsilon_n + g \langle n | \hat{U} \left(|n\rangle + \sum_{m \neq n} |m\rangle \frac{g}{E_n - \varepsilon_m} \langle m | \hat{U} | n \rangle \right)$$

$$= \varepsilon_n + g \langle n | \hat{U} | n \rangle + g^2 \sum_{m \neq n} \frac{\left| \langle m | \hat{U} | n \rangle \right|^2}{E_n - \varepsilon_m}$$

which is the second-order energy.

So the BWPT and the RSPT agree at each order of perturbation theory, as they must.

The structure of the equations, however, is very different.

A simple example shows the very different properties of the two methods.

Consider the Hamiltonian given by $\hat{H} = \hat{H}_0 + \hat{V}$ where

$$\hat{H}_0 = \begin{pmatrix} \varepsilon_1 & 0 \\ 0 & \varepsilon_2 \end{pmatrix} \rightarrow \text{eigenvectors } |1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{and } |2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and

$$\hat{V} = \begin{pmatrix} 0 & \alpha \\ \alpha^* & 0 \end{pmatrix}$$

The exact energy eigenvalues are obtained by diagonalizing the $\hat{H}$ matrix

$$\hat{H} = \begin{pmatrix} \varepsilon_1 & \alpha \\ \alpha^* & \varepsilon_2 \end{pmatrix}$$

to get the characteristic equation

$$\det \begin{bmatrix} \varepsilon_1 - E & \alpha \\ \alpha^* & \varepsilon_2 - E \end{bmatrix} = 0 = E^2 - (\varepsilon_1 + \varepsilon_2)E + (\varepsilon_1\varepsilon_2 - |\alpha|^2)$$

which has solutions

$$E_1 = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) + \frac{1}{2}\sqrt{(\varepsilon_1 - \varepsilon_2)^2 + 4|\alpha|^2}$$

$$E_2 = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{1}{2}\sqrt{(\varepsilon_1 - \varepsilon_2)^2 + 4|\alpha|^2}$$

In the *degenerate limit*, $\varepsilon_1 = \varepsilon_2 = \varepsilon$, we have the *exact* solutions

$$E_1 = \varepsilon + |\alpha| \text{ and } E_2 = \varepsilon - |\alpha|$$

Now, BWPT gives

$$E_n = \varepsilon_n + \langle n| \hat{V} |N\rangle$$

$$= \varepsilon_n + \langle n| \hat{V} |n\rangle + \sum_{m \neq n} \frac{\left|\langle m| \hat{V} |n\rangle\right|^2}{E_n - \varepsilon_m}$$

or

$$E_1 = \varepsilon_n + \langle 1| \hat{V} |1\rangle + \frac{\left|\langle 2| \hat{V} |1\rangle\right|^2}{E_1 - \varepsilon_2} = \varepsilon_1 + \frac{|\alpha|^2}{E_1 - \varepsilon_2}$$

Rearranging we get

$$E_1^2 - (\varepsilon_1 + \varepsilon_2)E_1 + (\varepsilon_1\varepsilon_2 - |\alpha|^2) = 0$$

which is the same eigenvalue equation as the exact solution.

In the degenerate limit, $\varepsilon_1 = \varepsilon_2 = \varepsilon$, we have

$$E_1 = \varepsilon + |\alpha|$$

or BWPT gives the exact answer for this simple system, even in the degenerate case.

On the other hand, RSPT gives to second-order

$$E_1 = \varepsilon_1 + \langle 1| \hat{V} |1\rangle + \frac{|\alpha|^2}{\varepsilon_1 - \varepsilon_2}$$

This is equivalent to the exact formula to second order only!

In the degenerate limit we get nonsense since the denominator vanishes.

As we saw earlier, RSPT requires an entirely different procedure in the degenerate case.

Notice that RSPT is in trouble even if $\varepsilon_1 \approx \varepsilon_2$ which implies that

$$\frac{|\alpha|^2}{\varepsilon_1 - \varepsilon_2} \text{ is very large}$$

and thus, that the perturbation expansion makes no sense(the terms are supposed to get smaller!).

A clever trick for handling these almost degenerate cases using RSPT goes as follows.

Almost Degenerate Perturbation Theory

Given $\hat{H} = \hat{H}_0 + \hat{V}$, suppose, as in the last example, we have two states $|n\rangle$ and $|m\rangle$ of the unperturbed (zero order) Hamiltonian $\hat{H}_0$ that have energies that are approximately equal.

This is a troublesome situation for RSPT because it is an expansion that includes increasing numbers of

$$\frac{1}{\varepsilon_n - \varepsilon_m}$$

terms.

This implies that successive terms in the perturbation series might decrease slowly or not at all.

To develop a more rapidly converging perturbation expansion we rearrange the calculation as follows.

We use the definition of the identity operator in terms of projection operators to write

$$\hat{V} = \hat{I}\hat{V}\hat{I} = \sum_{i,j} |i\rangle \langle i| \hat{V} |j\rangle \langle j|$$

We then break up $\hat{V}$ into two parts

$$\hat{V} = \hat{V}_1 + \hat{V}_2$$

where we separate out the m and n terms into $\hat{V}_1$

$$\hat{V}_1 = |m\rangle\langle m|\hat{V}|m\rangle\langle m| + |m\rangle\langle m|\hat{V}|n\rangle\langle n| + |n\rangle\langle n|\hat{V}|m\rangle\langle m| + |n\rangle\langle n|\hat{V}|n\rangle\langle n|$$

and $\hat{V}_2$ = the rest of the terms.

We then write

$$\hat{H} = \hat{H}_0 + \hat{V} = \hat{H}_0 + \hat{V}_1 + \hat{V}_2 = \hat{H}_0' + \hat{V}_2$$

This new procedure then finds exact eigenvectors/eigenvalues of $\hat{H}_0'$ and treats $\hat{V}_2$ by ordinary perturbation theory.

Since the basis is orthonormal, we have from the definition of $\hat{V}_2$, i.e.,

$$\hat{V}_2 = \sum_{i,j\neq m,n} |i\rangle\langle i|\hat{V}|j\rangle\langle j|$$

which gives

$$0 = \langle n|\hat{V}_2|n\rangle = \langle n|\hat{V}_2|m\rangle = \langle m|\hat{V}_2|n\rangle = \langle m|\hat{V}_2|m\rangle$$

Thus, the closeness of the levels ε_n and ε_m will not prevent us from applying standard perturbation theory to $\hat{V}_2$, i.e., the numerators of terms with very small energy denominators, which might cause the series to diverge, all vanish identically!

Now if $|i\rangle$ is an eigenvector of $\hat{H}_0$ (not $|m\rangle$ or $|n\rangle$), then it is also an eigenvector of $\hat{H}_0'$ since, by the orthonormality condition,

$$\hat{V}_1|i\rangle = 0$$

Neither $|m\rangle$ nor $|n\rangle$ is an eigenvector of $\hat{H}_0'$ however.

Now, the $\hat{H}_0$ matrix is diagonal since we are using its eigenvectors as a basis.

The $\hat{H}_0'$ matrix is diagonal also except for the 2×2 submatrix

$$\begin{pmatrix} \langle m|\hat{H}_0'|m\rangle & \langle m|\hat{H}_0'|n\rangle \\ \langle n|\hat{H}_0'|m\rangle & \langle n|\hat{H}_0'|n\rangle \end{pmatrix}$$

Therefore, we can finish the solution of the problem of $\hat{H}_0'$ by diagonalizing this 2×2 matrix.

Diagonalizing the 2×2 matrix is equivalent to finding the linear combinations (or new zero order eigenvectors)

$$\alpha|n\rangle + \beta|m\rangle$$

that diagonalize the 2×2 matrix.

We must have

$$\hat{H}_0'(\alpha|n\rangle + \beta|m\rangle) = \left(\hat{H}_0 + \hat{V}_1\right)(\alpha|n\rangle + \beta|m\rangle) = E'(\alpha|n\rangle + \beta|m\rangle)$$

Now

$$\begin{aligned}
\hat{H}_0'|n\rangle &= \hat{H}_0|n\rangle + \hat{V}_1|n\rangle \\
&= \varepsilon_n|n\rangle + |m\rangle\langle m|\hat{V}|m\rangle\langle m \mid n\rangle + |m\rangle\langle m|\hat{V}|n\rangle\langle n \mid n\rangle \\
&\qquad + |n\rangle\langle n|\hat{V}|m\rangle\langle m \mid n\rangle + |n\rangle\langle n|\hat{V}|n\rangle\langle n \mid n\rangle \\
&= \varepsilon_n|n\rangle + |m\rangle\langle m|\hat{V}|n\rangle + |n\rangle\langle n|\hat{V}|n\rangle \\
&= \left(\varepsilon_n + \langle n|\hat{V}|n\rangle\right)|n\rangle + |m\rangle\langle m|\hat{V}|n\rangle \\
&= E_n^{(1)}|n\rangle + \langle m|\hat{V}|n\rangle|m\rangle
\end{aligned}$$

and similarly

$$\hat{H}_0' |m\rangle = E_m^{(1)} |m\rangle + \langle n| \hat{V} |m\rangle |n\rangle$$

Therefore, we get

$$\alpha \left(E_n^{(1)} |n\rangle + \langle m| \hat{V} |n\rangle |m\rangle \right) + \beta \left(E_m^{(1)} |m\rangle + \langle n| \hat{V} |m\rangle |n\rangle \right) = E' \left(\alpha |n\rangle + \beta |m\rangle \right)$$

Since the state vectors $|m\rangle$ and $|n\rangle$ are orthogonal, we must then have

$$E_n^{(1)} \alpha + \langle n| \hat{V} |m\rangle \beta = E' \alpha$$

$$\langle m| \hat{V} |n\rangle \alpha + E_m^{(1)} \beta = E' \beta$$

These equations have two solutions, namely,

$$\alpha = \langle n| \hat{V} |m\rangle$$

$$\beta_{\pm} = \frac{E_m^{(1)} - E_n^{(1)}}{2} \pm \sqrt{\left(\frac{E_m^{(1)} - E_n^{(1)}}{2} \right)^2 + \left| \langle n| \hat{V} |m\rangle \right|^2}$$

which then give the results

$$E_{\pm} = \frac{E_m^{(1)} + E_n^{(1)}}{2} \pm \sqrt{\left(\frac{E_m^{(1)} - E_n^{(1)}}{2} \right)^2 + \left| \langle n| \hat{V} |m\rangle \right|^2}$$

We then know all the eigenvectors/eigenvalues of $\hat{H}_0'$ (we know all of the unpeturbed states) and we can deal with $\hat{V}_2$ by perturbation theory.

Finally, let us introduce another interesting idea.

Fake Degenerate Perturbation Theory

Consider the problem of finding the energy eigenvalues and state vectors for a system with a Hamiltonian $\hat{H} = \hat{H}_0 + \hat{V}$ where we know the solution to the zero-order system

$$\hat{H}_0 |n\rangle = \varepsilon_n |n\rangle$$

We will assume that the unperturbed states are nondegenerate.

Now define

$$E_{average} = E_{av} = \frac{1}{n} \sum_n \varepsilon_n$$

and redefine

$$\hat{H} = E_{av} \hat{I} + \hat{U}$$

where

$$U = H_0 - E_{av} I + V$$

If the energies associated with $\hat{U}$ are small corrections to E_{av}, then we can use degenerate perturbation theory to solve this problem, i.e., the new unperturbed Hamiltonian is

$$\hat{H}_0' = E_{av} \hat{I}$$

and all of its levels are degenerate in zero order

The problem is then solved by diagonalizing the $\hat{U}$ matrix in the basis of $\hat{H}_0$ states.

Thoughts on Degeneracy and Position Representation

When we derived the energy spectrum of the hydrogen atom we found that the states were labeled by three quantum numbers

$$|\psi\rangle = |n\ell m\rangle$$

where

$n =$ the radial quantum number

$\ell =$ orbital angular momentum quantum number

$m = z -$ component of orbital angular momentum quantum number

and we found that

$n = 1, 2, 3, \ldots\ldots\ldots$

$\ell = 0, 1, 2, \ldots\ldots, n-1$ for a given value of n

$m = -\ell, -\ell+1, \ldots\ldots \ell-1, \ell$ for a given value of ℓ

The energy eigenvalues, however, did not depend on ℓ or m.

We found that

$$E_{n\ell m} = E_n = -\frac{e^2}{2a_0 n^2}$$

Therefore, each energy level had a degeneracy given by

$$g = \sum_{\ell=0}^{n-1} \sum_{m=-\ell}^{\ell} 1 = \sum_{\ell=0}^{n-1} (2\ell+1) = 2\sum_{\ell=0}^{n-1} \ell + \sum_{\ell=0}^{n-1} 1$$

$$= 2\frac{n(n-1)}{2} + n = n^2$$

The degeneracy with respect to *m* is understandable since no direction is explicitly preferred in the Hamiltonian.

We expect that this degeneracy will disappear as soon as a preferred direction is added to the Hamiltonian, as in the case of external electric(Stark effect) or magnetic(Zeeman effect) fields.

The degeneracy with respect to ℓ is a property peculiar to the pure 1/r Coulomb potential.

Since no other atom except hydrogen has a pure Coulomb potential, we expect this degeneracy to vanish in other atoms.

Such a degeneracy is called an *accidental degeneracy*.

Now the electron and proton making up the hydrogen atom also have spin angular momentum.

The presence of these extra(internal) degrees of freedom should change the Hamiltonian.

The Schrodinger equation was derived from the eigenvalue equation for the Hamiltonian

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

by re-expressing that equation in the position representation.

The associated Schrodinger wave functions were given by the scalar product(linear functional) relation

$$\psi(\vec{r}) = \langle \vec{r} \mid \psi \rangle$$

The single particle Schrodinger equation is relevant for problems where the Hamiltonian contains terms dependent on ordinary 3–dimensional space(for many-particle systems we must use a multi-dimensional configuration space which bears no simple relationship to ordinary three-dimensional space).

Spin is an internal degree of freedom that has no representation in the 3–dimensional space of the Schrodinger wave equation.

The Schrodinger picture, however, does not choose a particular representation and, therefore, we can include spin within the context of solving the Schrodinger equation in the following ad hoc manner.

A more rigorous treatment requires relativity.

If there are spin-dependent terms in the Hamiltonian, then we expand the Hilbert space used to solved the problem by constructing a new basis that is made up of direct product states of the following type

$$|\psi_{new}\rangle = |\psi\rangle \otimes |s, m_s\rangle$$

where $|\psi\rangle$ depends on only ordinary 3–dimensional space and $|s, m_s\rangle$ is an eigenvector of $\vec{S}^2$ and $\hat{S}_z$.

The energy eigenvalue equation becomes

$$\begin{aligned} \hat{H}\,|\psi_{new}\rangle &= E\,|\psi_{new}\rangle \\ &= (((3-\text{space operators}))\,|\psi\rangle) \otimes (((\text{spin - dependent operators}))\,|s, m_s\rangle) \end{aligned}$$

and the corresponding wave function is

$$\langle \vec{r} \mid \psi_{new}\rangle = \langle \vec{r} \mid \psi\rangle\,|s, m_s\rangle = \psi(\vec{r})\,|s, m_s\rangle$$

where abstract spin vector is stuck onto the wave function in some way (maybe with superglue).

Let us now investigate what happens in atoms when we add in spin, some aspects of relativity and external fields.

We restrict our attention to one-electron atoms like hydrogen at this point.

Spin-Orbit Interaction - Fine Structure

The proton in hydrogen generates an electric field

$$\vec{\mathcal{E}} = \frac{e}{r^2}\hat{r} = \frac{e}{r^3}\vec{r}$$

that acts on the moving electron.

This result is approximately true (to first order) in most atoms.

Now special relativity says that an electron moving with a velocity $\vec{v}$ through an electric field $\vec{\mathcal{E}}$ also behaves as if it is interacting with a magnetic field given by

$$\vec{B} = -\frac{1}{c}\vec{v} \times \vec{\mathcal{E}}$$

to first order in v/c.

This magnetic field interacts with the spin (actually with its associated magnetic moment) to produce an additional contribution to the energy of the form

$$E = -\vec{M}_{spin} \cdot \vec{B}$$

where

$$\vec{M}_{spin} = -\frac{e}{mc}\vec{S}$$

Substituting everything in we get

$$E = -\frac{e}{mc^2}\vec{S} \cdot \left(\vec{v} \times \vec{\mathcal{E}}\right) = -\frac{e}{mc^2}\vec{S} \cdot \left(\vec{v} \times \frac{e}{r^3}\vec{r}\right)$$

$$= \frac{1}{m^2c^2}\vec{S} \cdot \vec{L}\frac{e^2}{r^3}$$

Now

$$\frac{e^2}{r^3} = \frac{1}{r}\frac{dV}{dr} \text{ for } V(r) = -\frac{e^2}{r} = \text{ potential energy of the electron}$$

so that we finally obtain the so-called spin-orbit energy contribution

$$E = \left[\frac{1}{m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}\cdot\vec{L} = E_{spin-orbit} = E_{so}$$

This corresponds to an additional term in the Hamiltonian of the form

$$\hat{H}_{so} = \left[\frac{1}{m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op}\cdot\vec{L}_{op}$$

This term couples the orbital and spin angular momentum degrees of freedom (hence the label spin-orbit energy) and mixes 3–dimensional space with spin space.

That is why we had to expand the Hilbert space as we discussed earlier.

Another way to think about this interaction is that the electron spin magnetic moment vector (or spin vector) is precessing about the direction of the magnetic field.

The equations for such a precessional motion are

$$\vec{M}_{spin}\times\vec{B} = \frac{d\vec{S}}{dt} = \vec{\Omega}_{L(armor)}\times\vec{S}$$

where

$$\left|\vec{\Omega}_L\right| = \frac{eB}{m}$$

Now

$$\vec{B} = -\frac{1}{c}\vec{v}\times\vec{\mathcal{E}} = \frac{1}{emc^2}\frac{1}{r}\frac{dV}{dr}\vec{L}$$

which implies that

$$\left|\vec{\Omega}_L\right| = \frac{1}{m^2c^2}\frac{1}{r}\frac{dV}{dr}\left|\vec{L}\right|$$

It turns out that this is exactly a factor of 2 too large.

There is another relativistic effect, which gives another precession (called Thomas precession) effect, that cancels exactly one-half of this spin-orbit effect.

Thomas Precession

This is a relativistic kinematic effect.

It results from the time dilation between the rest frames of the electron and the proton.

This causes observers in these two frames to disagree on the time required for one of the particles to a make a complete revolution about the other particle.

If an observer on the electron measures a time interval T , then the observer on the proton measures

$$T' = \gamma T \text{ where } \gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \quad , \quad v = \text{ speed of the electron}$$

We assume uniform circular motion for simplicity.

The orbital angular velocities measured by the observers are

$$\frac{2\pi}{T} \text{ and } \frac{2\pi}{T'}$$

respectively.

In the rest frame of the electron, the spin angular momentum vector maintains its direction in space.

This implies that an observer on the proton sees this spin vector precessing at a rate equal to the *difference* of the two angular velocities, i.e., the precessional frequency is

$$\begin{aligned}\Omega_{Thomas} &= \frac{2\pi}{T} - \frac{2\pi}{T'} = \frac{2\pi}{T'}\left(\frac{T'}{T} - 1\right) \\ &= \frac{2\pi}{T'}\left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} - 1\right) \\ &\approx \frac{2\pi}{T'}\left(\frac{v^2}{2c^2}\right)\end{aligned}$$

But we also have

$$\frac{2\pi}{T'} = \omega = \frac{|\vec{L}|}{mr^2} \text{ and } \frac{mv^2}{r} = -\frac{dV}{dr}$$

for circular motion.

Thus, we get

$$|\vec{\Omega}_T| = -\frac{1}{2}\frac{1}{m^2c^2}\frac{1}{r}\frac{dV}{dr}|\vec{L}| = -\frac{1}{2}|\vec{\Omega}_L|$$

Therefore, the combined precession is reduced by a factor of two and we get the result

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op}\cdot\vec{L}_{op}$$

The energy levels arising from this correction are called the *atomic fine structure*.

Another Relativity Correction

The correct relativistic kinetic energy term is

$$K = (\gamma - 1)mc^2 = \left(\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} - 1 \right) mc^2$$

$$= \left(\left(1 + \frac{1}{2}\frac{v^2}{c^2} - \frac{1}{8}\frac{v^4}{c^4} + \ldots \right) - 1 \right) mc^2$$

$$= \left(\frac{1}{2}\frac{v^2}{c^2} - \frac{1}{8}\frac{v^4}{c^4} \right) mc^2 = \frac{\vec{p}_{op}^{\,2}}{2m} - \frac{\vec{p}_{op}^{\,4}}{8m^3c^2}$$

Therefore, we must correct the $\vec{p}^2_{op}/2m$ we have already included in the Hamiltonian by adding a terms of the form

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^{\,4}}{8m^3c^2}$$

Thus, if no external field are present we have the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}_{relativity} + \hat{H}_{so}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{op}^{\,2}}{2m} - e^2 \left(\frac{1}{r} \right)_{op}$$

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^{\,4}}{8m^3c^2}$$

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2} \frac{1}{r} \frac{dV}{dr} \right] \vec{S}_{op} \cdot \vec{L}_{op}$$

External Fields - Zeeman and Stark Effects; Hyperfine Structure

Zeeman Effect

If an external magnetic field exists, then it interacts with the total magnetic moment of the electron, where

$$\vec{M}_{total} = \vec{M}_{orbital} + \vec{M}_{spin} = -\frac{e}{2mc}\left(g_\ell \vec{L} + g_s \vec{S}\right)$$

as we derived earlier.

If we define

$$\mu_B = \text{ Bohr magneton } = \frac{e\hbar}{mc}$$

and let $\vec{B}_{ext} = B\ \hat{z}$, then we have, using $g_\ell = 1$ and $g_s = 2$, the result

$$E_{Zeeman} = -\vec{M}_{total} \cdot \vec{B}_{ext} = \frac{\mu_B B}{\hbar}(L_z + 2S_z)$$

Thus, we must add a term of the form

$$\hat{H}_{Zeeman} = \frac{\mu_B B}{\hbar}(\hat{L}_z + 2\hat{S}_z)$$

to the Hamiltonian when an external magnetic field is present.

We can see directly how the orbital angular momentum part of this energy arises.

We saw earlier that if we had a Hamiltonian

$$\hat{H}_0 = \frac{\vec{p}_{op}^{\,2}}{2m} + V(\vec{r}_{op})$$

and we add an electromagnetic field characterized by a vector potential $\vec{A}$, where $\vec{B} = \nabla \times \vec{A}$, then the momentum operator changes to

$$\vec{p}_{em} = \vec{p} - \frac{e}{c}\vec{A}(\vec{r})$$

and the Hamiltonian changes to

$$\hat{H} = \frac{\vec{p}_{em,op}^2}{2m} + V(\vec{r}_{op}) = \frac{\left(\vec{p}_{op} - \frac{e}{c}\vec{A}(\vec{r}_{op})\right)^2}{2m} + V(\vec{r}_{op})$$

$$= \hat{H}_0 - \frac{e}{2mc}\left(\vec{p}_{op}\cdot\vec{A}(\vec{r}_{op}) + \vec{A}(\vec{r}_{op})\cdot\vec{p}_{op}\right) + \frac{e^2}{2mc^2}\vec{A}^2(\vec{r}_{op})$$

The magnetic field has to be enormous or the radial quantum number *n* very large for the $\vec{A}^2$ term to have any effect, so we will neglect it for now.

Let us look at the term

$$\vec{p}_{op}\cdot\vec{A}(\vec{r}_{op}) + \vec{A}(\vec{r}_{op})\cdot\vec{p}_{op}$$

For a uniform (constant in magnitude and direction) external field $\vec{B}$, we have

$$\vec{A} = -\frac{1}{2}\vec{r}\times\vec{B}$$

I will prove this so we get a chance to see the use of ε_{ijk} in vector algebra.

$$\nabla \times \vec{A} = -\frac{1}{2}\nabla \times \left(\vec{r} \times \vec{B}\right) = -\frac{1}{2}\sum_{ijk} \varepsilon_{ijk} \frac{\partial}{\partial x_j} \left(\vec{r} \times \vec{B}\right)_k \hat{e}_i$$

$$= -\frac{1}{2}\sum_{ijk} \varepsilon_{ijk} \frac{\partial}{\partial x_j} \left(\sum_{mn} \varepsilon_{kmn} x_m B_n\right) \hat{e}_i$$

$$= -\frac{1}{2}\sum_{ij}\sum_{mn} \left(\sum_{k} \varepsilon_{ijk}\varepsilon_{mnk}\right) \frac{\partial}{\partial x_j} \left(x_m B_n\right) \hat{e}_i$$

$$= -\frac{1}{2}\sum_{ij}\sum_{mn} \left(\delta_{im}\delta_{jn} - \delta_{in}\delta_{jm}\right) \frac{\partial}{\partial x_j} \left(x_m B_n\right) \hat{e}_i$$

$$= -\frac{1}{2}\sum_{ij} \left[\frac{\partial}{\partial x_j} \left(x_i B_j\right) \hat{e}_i - \frac{\partial}{\partial x_j} \left(x_j B_i\right) \hat{e}_i\right]$$

$$= -\frac{1}{2}\sum_{ij} \left[\frac{\partial x_i}{\partial x_j} B_j + x_i \frac{\partial B_j}{\partial x_j} - \frac{\partial x_j}{\partial x_j} B_i - x_j \frac{\partial B_i}{\partial x_j}\right] \hat{e}_i$$

Now

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij} \text{ and } \frac{\partial B_i}{\partial x_j} = 0$$

so we get

$$\nabla \times \vec{A} = -\frac{1}{2}\sum_{ij} \left[\delta_{ij} B_j - \delta_{jj} B_i\right] \hat{e}_i$$

$$= -\frac{1}{2}\left[\sum_{i} B_i \hat{e}_i - 3\sum_{i} B_i \hat{e}_i\right]$$

$$= -\frac{1}{2}\left[\vec{B} - 3\vec{B}\right] = \vec{B}$$

Therefore, we have

$$\vec{p}_{op}\cdot\vec{A}(\vec{r}_{op})+\vec{A}(\vec{r}_{op})\cdot\vec{p}_{op}=-\frac{1}{2}\left[\vec{p}_{op}\cdot\left(\vec{r}_{op}\times\vec{B}\right)+\left(\vec{r}_{op}\times\vec{B}\right)\cdot\vec{p}_{op}\right]$$

$$=-\frac{1}{2}\left[\sum_{ijk}\varepsilon_{ijk}B_k\left[\hat{p}_i\hat{x}_j+\hat{x}_j\hat{p}_i\right]\right]=-\frac{1}{2}\left[\sum_{ijk}\varepsilon_{ijk}B_k\left[2\hat{x}_j\hat{p}_i-i\hbar\delta_{ij}\right]\right]$$

$$=\sum_{ijk}\varepsilon_{kji}\hat{x}_j\hat{p}_iB_k=(\vec{r}_{op}\times\vec{p}_{op})\cdot\vec{B}=\vec{L}_{op}\cdot\vec{B}$$

which then gives

$$\hat{H}=\hat{H}_0-\frac{e}{2mc}\vec{L}_{op}\cdot\vec{B}+\frac{e^2}{2mc^2}\vec{A}^2(\vec{r}_{op})$$

which accounts for the orbital angular momentum part of the Zeeman energy.

The spin angular momentum part of the Zeeman energy cannot be derived from the non-relativistic Schrodinger equation.

When one derives the Dirac relativistic equation for the electron, the $\vec{S}_{op}\cdot\vec{B}$ term appears naturally.

Stark Effect

If a hydrogen atom is placed in an external electric field $\vec{\mathcal{E}}$ which is constant is space and time (uniform and static), then an additional energy appears.

It corresponds to an interaction between and electric dipole made up of the electron and proton separated by a distance and the external electric field.

We introduce the electric dipole moment operator

$$\vec{d}_{op}=-e\vec{r}_{op}$$

where $\vec{r}$ is the position vector of the electron relative to the proton.

We then write the extra energy term as

$$\hat{H}_{dipole} = -\vec{d}_{op} \cdot \vec{\mathcal{E}}$$

If we choose $\vec{E} = E\,\hat{z}$, then we have $\hat{H}_{\text{dipole}} = -ez\,\mathcal{E}$.

The full Hamiltonian is then

$$\hat{H} = \hat{H}_0 + \hat{H}_{relativity} + \hat{H}_{so} + \hat{H}_{Zeeman} + \hat{H}_{dipole}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{op}^{\,2}}{2m} - e^2\left(\frac{1}{r}\right)_{op}$$

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^{\,4}}{8m^3c^2}$$

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op} \cdot \vec{L}_{op}$$

$$\hat{H}_{Zeeman} = \frac{\mu_B}{\hbar}(\vec{L}_{op} + 2\vec{S}_{op}) \cdot \vec{B}$$

$$\hat{H}_{dipole} = -e\vec{r}_{op} \cdot \vec{\mathcal{E}}$$

Hyperfine Structure

The nuclear magnetic dipole moment also generates a magnetic field.

If we assume that it is a point dipole $\vec{M}_N$, then the magnetic field is given by

$$\vec{B}(\vec{r}) = \left(\frac{3\left(\vec{M}_N \cdot \vec{r}\right)\vec{r}}{r^5} - \frac{\vec{M}_N}{r^3} \right) + \frac{8\pi}{3}\vec{M}_N \delta(\vec{r})$$

where the first two terms are the standard result of the magnetic field due to a loop of current as seen from very far away (approximates dipole as a point) and the last term is peculiar to a point dipole.

The last term will give a contribution only for spherically symmetric states ($\ell = 0$).

extra energy is then

$$\hat{H}_{hyperfine} = -\vec{M}_e \cdot \vec{B}$$

$$= -\left(\frac{3\left(\vec{M}_N \cdot \vec{r}\right)\left(\vec{M}_e \cdot \vec{r}\right)}{r^5} - \frac{\vec{M}_N \cdot \vec{M}_e}{r^3} \right) - \frac{8\pi}{3}\vec{M}_N \cdot \vec{M}_e \delta(\vec{r})$$

where

$$\vec{M}_N = g_N \frac{Ze}{2m_N c}\vec{S}_{N,op} \text{ and } \vec{M}_e = \frac{e}{mc}\vec{S}_{e,op}$$

This is clearly due to spin-spin interactions between the electron and the nucleus and gives rise to the so-called hyperfine level splitting.

Examples

Now that we have identified all of the relevant corrections to the Hamiltonian for atoms, let us illustrate the procedures for calculation of the new energy levels via perturbation theory.

We look at the simplest atom first.

Spin-Orbit, Relativity, Zeeman Effect in Hydrogen Atom

The Hamiltonian is

$$\hat{H} = \hat{H}_0 + \hat{H}_{relativity} + \hat{H}_{so} + \hat{H}_{Zeeman}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{op}^2}{2m} - e^2\left(\frac{1}{r}\right)_{op}$$

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^4}{8m^3c^2}$$

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op}\cdot\vec{L}_{op}$$

$$\hat{H}_{Zeeman} = \frac{\mu_B}{\hbar}(\vec{L}_{op} + 2\vec{S}_{op})\cdot\vec{B}$$

The first step is to calculate all relevant commutators so that we can find those operators that have a common eigenbasis.

$$\left[\vec{J}_{op}^2, \hat{H}_0\right] = 0 = \left[\hat{J}_z, \hat{H}_0\right] = \left[\vec{L}_{op}^2, \hat{H}_0\right] = \left[\vec{S}_{op}^2, \hat{H}_0\right]$$

$$\left[\vec{L}_{op}^2, \vec{S}_{op}^2\right] = 0 = \left[\vec{L}_{op}^2, \vec{J}_{op}^2\right] = \left[\vec{S}_{op}^2, \vec{J}_{op}^2\right]$$

$$\left[\vec{J}_{op}^2, \hat{J}_z\right] = 0 = \left[\vec{L}_{op}^2, \hat{J}_z\right] = \left[\vec{S}_{op}^2, \hat{J}_z\right]$$

This says that there exists a common set of eigenvectors in the unperturbed system for the set of operators

$$\hat{H}_0, \vec{J}_{op}^2, \vec{L}_{op}^2, \vec{S}_{op}^2, \hat{J}_z$$

We label these states by the corresponding eigenvalues of the commuting set of observables (these are called *good quantum numbers*)

$$|n, \ell, s, j, m_j\rangle$$

We also have

$$\left[\hat{S}_z, \hat{H}_0\right] = 0 = \left[\hat{L}_z, \hat{H}_0\right] = \left[\vec{L}^2_{op}, \hat{H}_0\right] = \left[\vec{S}^2_{op}, \hat{H}_0\right]$$
$$\left[\vec{L}^2_{op}, \vec{S}^2_{op}\right] = 0 = \left[\vec{L}^2_{op}, \hat{S}_z\right] = \left[\vec{S}^2_{op}, \hat{S}_z\right] = \left[\vec{L}^2_{op}, \hat{L}_z\right] = \left[\vec{S}^2_{op}, \hat{L}_z\right]$$

which says that there exists another common set of eigenvectors in the unperturbed system for the operators

$$\hat{H}_0, \vec{L}^2_{op}, \vec{S}^2_{op}, \hat{L}_z, \hat{S}_z$$

We label these states by the corresponding eigenvalues of this commuting set of observables (again these are called *good quantum numbers*)

$$|n, \ell, s, m_\ell, m_s\rangle$$

In this latter basis, the unperturbed or zero-order Hamiltonian has solutions represented by

$$\hat{H}_0 |n, \ell, m_\ell, s, m_s\rangle = E_n^{(0)} |n, \ell, m_\ell, s, m_s\rangle \quad , \quad E_n^{(0)} = -\frac{Ze^2}{2a_0 n^2}$$

$$Ze = \text{ nucleus charge } (Z = 1 \text{ for hydrogen }) \, , \; a_0 = \text{ Bohr radius } = \frac{\hbar^2}{me^2}$$

$$\psi_{n\ell m_\ell s m_s}(r, \theta, \phi) = \langle \vec{r}| \left(|nlm_\ell s m_s\rangle\right) = \langle \vec{r}| \left(|nlm_\ell\rangle |sm_s\rangle\right)$$
$$= \langle \vec{r} \mid nlm_\ell\rangle |sm_s\rangle = \psi_{n\ell m_\ell}(r, \theta, \phi) |sm_s\rangle$$

and first few unperturbed wave functions are

$$\psi_{100}(r,\theta,\phi) = \frac{1}{\sqrt{\pi}}\left(\frac{Z}{a_0}\right)^{3/2} e^{-\frac{Zr}{a_0}}$$

$$\psi_{200}(r,\theta,\phi) = \frac{1}{\sqrt{32\pi}}\left(\frac{Z}{a_0}\right)^{3/2}\left(2-\frac{Zr}{a_0}\right) e^{-\frac{Zr}{2a_0}}$$

$$\psi_{210}(r,\theta,\phi) = \frac{1}{\sqrt{32\pi}}\left(\frac{Z}{a_0}\right)^{3/2}\frac{Zr}{a_0} e^{-\frac{Zr}{2a_0}}\cos\theta$$

$$\psi_{21\pm1}(r,\theta,\phi) = \frac{1}{\sqrt{64\pi}}\left(\frac{Z}{a_0}\right)^{3/2}\frac{Zr}{a_0} e^{-\frac{Zr}{2a_0}}\sin\theta e^{\pm i\phi}$$

We also have the relations below for the unperturbed states.

$$\vec{L}^2_{op}\left|n,\ell,s,j,m_j\right\rangle = \hbar^2\ell(\ell+1)\left|n,\ell,s,j,m_j\right\rangle$$

$$\vec{S}^2_{op}\left|n,\ell,s,j,m_j\right\rangle = \hbar^2 s(s+1)\left|n,\ell,s,j,m_j\right\rangle$$

$$\vec{J}^2_{op}\left|n,\ell,s,j,m_j\right\rangle = \hbar j(j+1)\left|n,\ell,s,j,m_j\right\rangle$$

$$\hat{J}_z\left|n,\ell,s,j,m_j\right\rangle = \hbar^2 m_j\left|n,\ell,s,j,m_j\right\rangle$$

$$\vec{L}^2_{op}\left|n,\ell,s,m_\ell,m_s\right\rangle = \hbar^2\ell(\ell+1)\left|n,\ell,s,m_\ell,m_s\right\rangle$$

$$\vec{S}^2_{op}\left|n,\ell,s,m_\ell,m_s\right\rangle = \hbar^2 s(s+1)\left|n,\ell,s,m_\ell,m_s\right\rangle$$

$$\hat{L}_z\left|n,\ell,s,m_\ell,m_s\right\rangle = \hbar m_\ell\left|n,\ell,s,m_\ell,m_s\right\rangle$$

$$\hat{S}_z\left|n,\ell,s,m_\ell,m_s\right\rangle = \hbar m_s\left|n,\ell,s,m_\ell,m_s\right\rangle$$

$$\begin{aligned}\hat{J}_z\left|n,\ell,s,m_\ell,m_s\right\rangle &= (\hat{L}_z+\hat{S}_z)\left|n,\ell,s,m_\ell,m_s\right\rangle \\ &= \hbar(m_\ell+m_s)\left|n,\ell,s,m_\ell,m_s\right\rangle \\ &= \hbar m_j\left|n,\ell,s,m_\ell,m_s\right\rangle\end{aligned}$$

Since the total angular momentum is given by

$$\vec{J}_{op} = \vec{L}_{op} + \vec{S}_{op}$$

the rules we developed for the addition of angular momentum say that

$$j = \ell + s, \ell + s - 1, ..., |\ell - s| + 1, |\ell - s|$$

and

$$m_j = j, j - 1, j - 2, ..., -j + 1, -j$$

In the case of hydrogen, where $s = 1/2$, we have only two allowed total j values for each ℓ value, namely,

$$j = \ell \pm \frac{1}{2}$$

We can use either of the two sets of basis states (both are an orthonormal basis)

$$|n, \ell, s, j, m_j\rangle \text{ or } |nlm_\ell sm_s\rangle$$

as the zero-order states for a perturbation theory development of the energies.

The choice depends on the specific perturbations we are trying to calculate.

Let us start off by using the $|n, \ell, s, j, m_j\rangle$ states.

If we use the potential energy function $V(r) = -\frac{e^2}{r}$

for hydrogen, then the spin-orbit correction to the Hamiltonian becomes

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op} \cdot \vec{L}_{op} = \frac{e^2}{2m^2c^2}\left(\frac{1}{r^3}\right)\vec{S}_{op} \cdot \vec{L}_{op}$$

Now $\vec{J}_{op} = \vec{L}_{op} + \vec{S}_{op}$ implies that

$$\vec{J}_{op}^2 = \vec{L}_{op}^2 + \vec{S}_{op}^2 + 2\vec{L}_{op} \cdot \vec{S}_{op}$$

$$\rightarrow \vec{L}_{op} \cdot \vec{S}_{op} = \frac{1}{2}\left(\vec{J}_{op}^2 - \vec{L}_{op}^2 - \vec{S}_{op}^2\right)$$

and therefore

$$\hat{H}_{so} = \frac{e^2}{4m^2c^2}\left(\frac{1}{r^3}\right)\left(\vec{J}_{op}^2 - \vec{L}_{op}^2 - \vec{S}_{op}^2\right)$$

Therefore

$$\left[\vec{J}_{op}^2, \hat{H}_{so}\right] = 0 = \left[\hat{J}_z, \hat{H}_{so}\right] = \left[\vec{L}_{op}^2, \hat{H}_{so}\right] = \left[\vec{S}_{op}^2, \hat{H}_{so}\right] = \left[\hat{H}_0, \hat{H}_{so}\right]$$

which implies that the state vectors $|n, \ell, s, j, m_j\rangle$ are also eigenvectors of $\hat{H}_{so}$.

This means that the matrix representation of $\hat{H}_{so}$ in this basis will be diagonal and we can apply standard non-degenerate perturbation theory.

Applying our rules for first order perturbation theory we have

$$E_{n\ell sjm_j} = E_n^{(0)} + \langle n\ell sjm_j | \hat{H}_{so} | n\ell sjm_j \rangle$$

$$= E_n^{(0)} + \frac{e^2\hbar^2}{4m^2c^2}\left(j(j+1) - \ell(\ell+1) - s(s+1)\right)\langle n\ell sjm_j | \frac{1}{r_{op}^3} | n\ell sjm_j \rangle$$

We now evaluate

$$\langle n\ell sjm_j | \frac{1}{r_{op}^3} | n\ell sjm_j \rangle = \int d^3\vec{r}' \int d^3\vec{r}\, \langle n\ell sjm_j \mid \vec{r}' \rangle \langle \vec{r}' | \frac{1}{r_{op}^3} | \vec{r} \rangle \langle \vec{r} \mid n\ell sjm_j \rangle$$

Now

$$\langle \vec{r}' | \frac{1}{r_{op}^3} | \vec{r} \rangle = \frac{1}{r^3} \langle \vec{r}' \mid \vec{r} \rangle = \frac{1}{r^3} \delta(\vec{r}' - \vec{r})$$

which gives

$$\langle n\ell sjm_j | \frac{1}{r_{op}^3} | n\ell sjm_j \rangle = \int d^3\vec{r} \frac{1}{r^3} |\langle \vec{r} \mid n\ell sjm_j \rangle|^2 = \int d^3\vec{r} \frac{1}{r^3} |\psi_{n\ell sjm_j}(\vec{r})|^2$$

Therefore, we can calculate the energy corrections once we know $|n, \ell, s, j, m_j\rangle$.

We first consider the trivial case of the $n = 1$ level in hydrogen.

We have

$$E_1^{(0)} = -\frac{e^2}{2a_0}$$

and the corresponding states are shown in the Table below.

n	ℓ	s	m_ℓ	m_s	j	m_j
1	0	1/2	0	+1/2	1/2	+1/2
1	0	1/2	0	-1/2	1/2	-1/2

***n* = 1 level quantum numbers**

or

$$\left|1, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm} \text{ and } \left|1, 0, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right\rangle_{jm}$$

where we have added the label *jm* to distinguish them from the $|n, \ell, s, m_\ell, m_s\rangle$ states which we label with $m_\ell m_s = mm$.

We are able to specify m_ℓ and m_s also in this case because when $\ell = 0$ we must have $m_\ell = 0$ and $j = s$ which says that $m_j = m_s$.

This is a *two-fold degenerate* ground state for the atom in *zeroth order*.

Since $\ell = 0$, which implies that $j = s = 1/2$, the expectation value $\langle \hat{H}_{so} \rangle = 0$.

Thus, there is no spin orbit correction for this state to first order.

In fact, there is no spin orbit correction to any order for an $\ell = 0$ state.

Now in general, we can write

$$|n, \ell, s, j, m_j\rangle = \sum_{\substack{m_\ell, m_s \\ m_\ell + m_s = m_j}} a_{n\ell s j m_j m_\ell m_s} |n, \ell, s, m_\ell, m_s\rangle$$

where the $a_{n\ell s j m_j m_\ell m_s}$ are the relevant Clebsch-Gordon(CG) coefficients.

For the $n = 1$ level we have the simple cases where

$$\left|1, 0, \frac{1}{2}, \frac{1}{2}, \pm\frac{1}{2}\right\rangle_{jm} = \left|1, 0, \frac{1}{2}, 0, \pm\frac{1}{2}\right\rangle_{mm}$$

i.e., the CG coefficients are equal to 1,

$$a_{1,0,\frac{1}{2},\frac{1}{2},\pm\frac{1}{2},0,\pm\frac{1}{2}} = 1$$

which is always true for the (maximum j, maximum(minimum) m_j) states.

There is always only one such state.

The next level is the $n = 2$ level of hydrogen and the complexity of the calculation increases fast.

We have

$$E_2^{(0)} = -\frac{e^2}{8a_0}$$

It is always the case that the direct-product states $|n, \ell, s, m_\ell, m_s\rangle$ are easier to write down.

For this level the $|n, \ell, s, j, m_j\rangle$ states need to be constructed from the $|n, \ell, s, m_\ell, m_s\rangle$ states.

Before we proceed, we can enumerate the states in both schemes.

The degeneracy is given by

$$\text{degeneracy} = 2\sum_{\ell=0}^{n-1}(2\ell+1) = 2n^2 = 8 = \sum_{\ell=0}^{n-1}\sum_{j=|\ell-s|}^{\ell+s}(2j+1) = 2\sum_{\ell=0}^{n-1}(2\ell+1)$$

The states are shown in the Tables below.

n	ℓ	s	m_ℓ	m_s	ket
2	1	1/2	1	+1/2	$\|2,1,1/2,1,1/2\rangle_{mm}$
2	1	1/2	0	+1/2	$\|2,1,1/2,0,1/2\rangle_{mm}$
2	1	1/2	-1	+1/2	$\|2,1,1/2,-1,1/2\rangle_{mm}$
2	1	1/2	1	-1/2	$\|2,1,1/2,1,-1/2\rangle_{mm}$
2	1	1/2	0	-1/2	$\|2,1,1/2,0,-1/2\rangle_{mm}$
2	1	1/2	-1	-1/2	$\|2,1,1/2,-1,-1/2\rangle_{mm}$
2	0	1/2	0	1/2	$\|2,0,1/2,0,1/2\rangle_{mm}$
2	0	1/2	0	-1/2	$\|2,0,1/2,0,-1/2\rangle_{mm}$

: $n = 2$ level quantum numbers $m_\ell m_s$ states

n	ℓ	s	j	m_j	ket
2	1	1/2	3/2	3/2	$\|2,1,1/2,1,1/2\rangle_{mm}$
2	1	1/2	3/2	1/2	$\|2,1,1/2,0,1/2\rangle_{mm}$
2	1	1/2	3/2	-1/2	$\|2,1,1/2,-1,1/2\rangle_{mm}$
2	1	1/2	3/2	-3/2	$\|2,1,1/2,1,-1/2\rangle_{mm}$
2	1	1/2	1/2	1/2	$\|2,1,1/2,0,-1/2\rangle_{mm}$
2	1	1/2	1/2	-1/2	$\|2,1,1/2,-1,-1/2\rangle_{mm}$
2	0	1/2	1/2	1/2	$\|2,0,1/2,0,1/2\rangle_{mm}$
2	0	1/2	1/2	-1/2	$\|2,0,1/2,0,-1/2\rangle_{mm}$

$n = 2$ level quantum numbers jm_j states

In the first set of states, we could have also included the m_j label since we must have $m_\text{j} = m_\ell + m_\text{s}$.

In order to learn all the intricate details of this type of calculation, we shall proceed in two ways using the spin-orbit correction as an example.

In method #1, we will construct the $|n, \ell, s, j, m_j\rangle$ states (the zero-order state vectors) from the $|n, \ell, s, m_\ell, m_s\rangle$ and then calculate the first-order energy corrections. In this basis, the $\langle \hat{H}_{so} \rangle$ matrix will be diagonal.

In method #2, we will construct the $\langle \hat{H}_{so} \rangle$ matrix using the easiest states to write down, namely the $|n, \ell, s, m_\ell, m_s\rangle$ states, and then diagonalize it to find the correct first order energies and new zero-order state vectors, which should be the $|n, \ell, s, j, m_j\rangle$ states.

Method #1

We start with the state with maximum j and m_j values.

This state always has a CG coefficient equal to 1, i.e., there is only one way to construct it from the other angular momenta.

$$\left|n=2, \ell=1, s=\frac{1}{2}, j=\frac{3}{2}, m_j=\frac{3}{2}\right\rangle_{jm} = \left|n=2, \ell=1, s=\frac{1}{2}, m_\ell=1, m_s=\frac{1}{2}\right\rangle_{mm}$$

where we have shown all the labels explicitly.

From now on we will write such equations as

$$\left|2, 1, \frac{1}{2}, \frac{3}{2}, \frac{3}{2}\right\rangle_{jm} = \left|2, 1, \frac{1}{2}, 1, \frac{1}{2}\right\rangle_{mm}$$

We then use the lowering operators to obtain

$$\hat{J}_- \left|2,1,\frac{1}{2},\frac{3}{2},\frac{3}{2}\right\rangle_{jm} = \hbar\sqrt{\frac{3}{2}\left(\frac{3}{2}+1\right)-\frac{3}{2}\left(\frac{3}{2}-1\right)}\left|2,1,\frac{1}{2},\frac{3}{2},\frac{1}{2}\right\rangle_{jm}$$

$$= \hbar\sqrt{3}\left|2,1,\frac{1}{2},\frac{3}{2},\frac{1}{2}\right\rangle_{jm} = \left(\hat{L}_- + \hat{S}_-\right)\left|2,1,\frac{1}{2},1,\frac{1}{2}\right\rangle_{mm}$$

$$= \hbar\sqrt{1(1+1)-1(1-1)}\left|2,1,\frac{1}{2},0,\frac{1}{2}\right\rangle_{mm}$$

$$+ \hbar\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)-\frac{1}{2}\left(\frac{1}{2}-1\right)}\left|2,1,\frac{1}{2},1,-\frac{1}{2}\right\rangle_{mm}$$

$$= \hbar\sqrt{2}\left|2,1,\frac{1}{2},0,\frac{1}{2}\right\rangle_{mm} + \hbar\sqrt{1}\left|2,1,\frac{1}{2},1,-\frac{1}{2}\right\rangle_{mm}$$

or

$$\left|2,1,\frac{1}{2},\frac{3}{2},\frac{1}{2}\right\rangle_{jm} = \sqrt{\frac{2}{3}}\left|2,1,\frac{1}{2},0,\frac{1}{2}\right\rangle_{mm} + \sqrt{\frac{1}{3}}\left|2,1,\frac{1}{2},1,-\frac{1}{2}\right\rangle_{mm}$$

Notice that we use the total *J* operators on the left and the *L* and *S* operators on the right.

The result is a linear combination of the $|n,\ell,s,m_\ell,m_s\rangle$ states all with $m_j = 1/2$ as we expected.

Continuing this process we have

$$\hat{J}_- \left|2,1,\frac{1}{2},\frac{3}{2},\frac{1}{2}\right\rangle_{jm} = \hbar\sqrt{\frac{3}{2}\left(\frac{3}{2}+1\right)-\frac{1}{2}\left(\frac{1}{2}-1\right)}\left|2,1,\frac{1}{2},\frac{3}{2},-\frac{1}{2}\right\rangle_{jm}$$

$$= 2\hbar\left|2,1,\frac{1}{2},\frac{3}{2},-\frac{1}{2}\right\rangle_{jm}$$

$$= \left(\hat{L}_- + \hat{S}_-\right)\left(\sqrt{\frac{2}{3}}\left|2,1,\frac{1}{2},0,\frac{1}{2}\right\rangle_{mm} + \sqrt{\frac{1}{3}}\left|2,1,\frac{1}{2},1,-\frac{1}{2}\right\rangle_{mm}\right)$$

$$= \hbar\sqrt{\frac{4}{3}}\left|2,1,\frac{1}{2},-1,\frac{1}{2}\right\rangle_{mm} + \hbar\sqrt{\frac{8}{3}}\left|2,1,\frac{1}{2},0,-\frac{1}{2}\right\rangle_{mm}$$

or

$$\left|2,1,\frac{1}{2},\frac{3}{2},-\frac{1}{2}\right\rangle_{jm}$$

$$= \sqrt{\frac{1}{3}}\left|2,1,\frac{1}{2},-1,\frac{1}{2}\right\rangle_{mm} + \sqrt{\frac{2}{3}}\left|2,1,\frac{1}{2},0,-\frac{1}{2}\right\rangle_{mm}$$

and finally

$$\left|2,1,\frac{1}{2},\frac{3}{2},-\frac{3}{2}\right\rangle_{jm} = \left|2,1,\frac{1}{2},-1,-\frac{1}{2}\right\rangle_{mm}$$

We now need to construct the maximum state for then next lowest value of j, namely,

$$\left|2,1,\frac{1}{2},\frac{1}{2},\frac{1}{2}\right\rangle_{jm}$$

This state has $m_j = 1/2$ so it must be constructed out of the same states that make up

$$\left|2,1,\frac{1}{2},\frac{3}{2},\frac{1}{2}\right\rangle_{jm}$$

or it can be written as

$$\left|2,1,\frac{1}{2},\frac{1}{2},\frac{1}{2}\right\rangle_{jm} = a\left|2,1,\frac{1}{2},0,\frac{1}{2}\right\rangle_{mm} + b\left|2,1,\frac{1}{2},1,-\frac{1}{2}\right\rangle_{mm}$$

Now we must have

$$\left\langle 2,1,\frac{1}{2},\frac{3}{2},\frac{1}{2}\,\middle|\,2,1,\frac{1}{2},\frac{1}{2},\frac{1}{2}\right\rangle_{jm} = 0 \text{ orthogonality}$$

$$\left\langle 2,1,\frac{1}{2},\frac{1}{2},\frac{1}{2}\,\middle|\,2,1,\frac{1}{2},\frac{1}{2},\frac{1}{2}\right\rangle_{jm} = 1 \text{ normalization}$$

Using the orthonormality of the $|n,\ell,s,m_\ell,m_s\rangle$ states we get

$$\sqrt{\frac{2}{3}}a + \sqrt{\frac{1}{3}}b = 0 \text{ and } a^2 + b^2 = 1$$

The solution is

$$a = \sqrt{\frac{1}{3}} \qquad \text{and} \qquad b = -\sqrt{\frac{2}{3}}$$

and therefore

$$\left|2,1,\frac{1}{2},\frac{1}{2},\frac{1}{2}\right\rangle_{jm} = \sqrt{\frac{1}{3}}\left|2,1,\frac{1}{2},0,\frac{1}{2}\right\rangle_{mm} - \sqrt{\frac{2}{3}}\left|2,1,\frac{1}{2},1,-\frac{1}{2}\right\rangle_{mm}$$

In a similar manner, we find

$$\left|2,1,\frac{1}{2},\frac{1}{2},-\frac{1}{2}\right\rangle_{jm}$$

$$= \sqrt{\frac{2}{3}}\left|2,1,\frac{1}{2},-1,\frac{1}{2}\right\rangle_{mm} - \sqrt{\frac{1}{3}}\left|2,1,\frac{1}{2},0,-\frac{1}{2}\right\rangle_{mm}$$

Finally, we construct the other $j = 1/2$ states with $\ell = 0$.

They are

$$\left|2,0,\frac{1}{2},0,\frac{1}{2}\right\rangle_{jm} = \left|2,0,\frac{1}{2},0,\frac{1}{2}\right\rangle_{mm}$$

$$\left|2,0,\frac{1}{2},0,-\frac{1}{2}\right\rangle_{jm} = \left|2,0,\frac{1}{2},0,-\frac{1}{2}\right\rangle_{mm}$$

We can now calculate the first-order energy corrections.

We do not actually need the detailed construction of the states to do this, but we will need these states to compare with the results of Method #2 later.

We found earlier that in the $|n,\ell,s,j,m_j\rangle$ basis

$$E_{n\ell sjm_j} = E_n^{(0)} + \frac{e^2\hbar^2 A_{n\ell sjm_j}}{4m^2c^2}\left(j(j+1) - \ell(\ell+1) - s(s+1)\right)$$

$$A_{n\ell sjm_j} = \int d^3\vec{r}\,\frac{1}{r^3}\left|\psi_{n\ell sjm_j}(\vec{r})\right|^2$$

or

$$\Delta E_{n,\ell,s,j,m_j} = E_{n,\ell,s,j,m_j} - E_n^{(0)} = \frac{e^2\hbar^2 A_{\pm}}{4m^2c^2}\begin{cases}\ell & j = \ell + 1/2\\ -(\ell+1) & j = \ell - 1/2\\ 0 & \ell = 0\end{cases}$$

$$A_{\pm} = \int d^3\vec{r}\,\frac{1}{r^3}\left|\psi_{n,\ell,s,j=\ell\pm\frac{1}{2},m_j}(\vec{r})\right|^2$$

Evaluating the integrals we get

$$\Delta E_{n,\ell,s,j,m_j} = \frac{Z^2 \left|E_n^{(0)}\right| \alpha^2}{n(2\ell+1)(\ell+1)} \qquad j = \ell + \frac{1}{2}$$

$$\Delta E_{n,\ell,s,j,m_j} = -\frac{Z^2 \left|E_n^{(0)}\right| \alpha^2}{n\ell(2\ell+1)} \qquad j = \ell - \frac{1}{2}$$

$$\Delta E_{n,\ell,s,j,m_j} = 0 \qquad \ell = 0$$

where

$$\alpha = \frac{e^2}{\hbar c} = \text{ fine structure constant}$$

Therefore, for the $n = 2$ level we have

$$\Delta E_{2,1,\frac{1}{2},\frac{3}{2},m_j} = \frac{Z^2 \left|E_2^{(0)}\right| \alpha^2}{12} \qquad j = \ell + \frac{1}{2} = \frac{3}{2}$$

$$\Delta E_{2,1,\frac{1}{2},\frac{1}{2},m_j} = -\frac{Z^2 \left|E_2^{(0)}\right| \alpha^2}{6} \qquad j = \ell - \frac{1}{2} = \frac{1}{2}$$

$$\Delta E_{2,0,\frac{1}{2},\frac{1}{2},m_j} = 0 \qquad j = \ell + \frac{1}{2} = \frac{1}{2}$$

We note that for hydrogen $Z^2\alpha^2 \approx 10^{-4}$ and thus, the fine structure splitting is significantly smaller than the zero-order energies.

The relativity correction is the same order of magnitude as the spin-orbit correction.

We found

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^{\,4}}{8m^3c^2}$$

This gives the correction

$$\Delta E_{rel} = -\frac{\hbar^4}{8m^3c^2}\int_0^\infty r^2 dr \psi^*_{n\ell m}(r)\nabla^4\psi_{n\ell m}(r)$$

$$= \frac{Z^2\left|E_n^{(0)}\right|\alpha^2}{4n^2}\left(3-\frac{4n}{\ell+\frac{1}{2}}\right)$$

Combining these two correction terms(spin-orbit and relativity) gives

$$\Delta E_{fine\ structure} = \frac{Z^2\left|E_2^{(0)}\right|\alpha^2}{4n^2}\left(3-\frac{4n}{j+\frac{1}{2}}\right)$$

The result is independent of ℓ .

It turns out that this result is valid for $\ell = 0$ also.

There is an additional term that must be added to the Hamiltonian which contributes only in $\ell = 0$ states.

It is called the Darwin term.

The Darwin term comes from the relativistic equation for the electron and takes the form

$$\hat{H}_{Darwin} = \frac{\hbar^2}{8m^2c^2}\nabla^2 V = -\frac{\hbar^2}{8m^2c^2}\left(4\pi e Q_{nuclear}(\vec{r})\right) = \frac{\pi\hbar^2 Ze^2}{2m^2c^2}\delta(\vec{r})$$

where

$$Q_{nuclear}(\vec{r}) = \text{ the nuclear charge density}$$

Because of the delta function, a contribution

$$\left\langle \hat{H}_{Darwin} \right\rangle_{nj\ell} = \frac{\pi\hbar^2 Ze^2}{2m^2c^2} \left|\psi_{n\ell}(0)\right|^2 = \frac{mc^2 Z^4\alpha^4}{2n^3}\delta_{\ell,0}$$

arises for $\ell = 0$ states only.

This is identical to the contribution $\langle \hat{H}_{\text{so}} + \hat{H}_{\text{rel}} \rangle$ for $\ell = 0, j = 1/2$.

Method #2

We use the $|n, \ell, s, m_\ell, m_s\rangle$ basis.

In this case, the best form of the operators to use (means we know how to evaluate them with these states) are

$$\hat{H}_{so} = \frac{e^2}{2m^2c^2}\left(\frac{1}{r^3}\right)\vec{S}_{op}\cdot\vec{L}_{op}$$

$$= \frac{e^2}{2m^2c^2}\left(\frac{1}{r^3}\right)\left(\hat{L}_z\hat{S}_z + \frac{1}{2}\left(\hat{L}_+\hat{S}_- + \hat{L}_-\hat{S}_+\right)\right)$$

If we label the rows and columns of the matrix representation by

$$|1\rangle = \left|2, 1, \frac{1}{2}, 1, \frac{1}{2}\right\rangle_{mm} \qquad m_j = \frac{3}{2}$$

$$|2\rangle = \left|2, 1, \frac{1}{2}, 1, -\frac{1}{2}\right\rangle_{mm} \qquad m_j = \frac{1}{2}$$

$$|3\rangle = \left|2, 1, \frac{1}{2}, 0, \frac{1}{2}\right\rangle_{mm} \qquad m_j = \frac{1}{2}$$

$$|4\rangle = \left|2, 1, \frac{1}{2}, 0, -\frac{1}{2}\right\rangle_{mm} \qquad m_j = -\frac{1}{2}$$

$$|5\rangle = \left|2, 1, \frac{1}{2}, -1, \frac{1}{2}\right\rangle_{mm} \qquad m_j = -\frac{1}{2}$$

$$|6\rangle = \left|2, 1, \frac{1}{2}, -1, -\frac{1}{2}\right\rangle_{mm} \qquad m_j = -\frac{3}{2}$$

$$|7\rangle = \left|2, 0, \frac{1}{2}, 0, \frac{1}{2}\right\rangle_{mm} \qquad m_j = \frac{1}{2}$$

$$|8\rangle = \left|2, 0, \frac{1}{2}, 0, -\frac{1}{2}\right\rangle_{mm} \qquad m_j = -\frac{1}{2}$$

then we get the matrix for $\langle\ \hat{H}_{so}\ \rangle$ as shown in the Table below.

I have used a table rather than an equation format so that I could clearly label the rows and columns by the state index.

	1	2	3	4	5	6	7	8
1	a	0	0	0	0	0	0	0
2	0	b	c	0	0	0	0	0
3	0	c	d	0	0	0	0	0
4	0	0	0	e	f	0	0	0
5	0	0	0	f	g	0	0	0
6	0	0	0	0	0	h	0	0
7	0	0	0	0	0	0	p	0
8	0	0	0	0	0	0	0	q

$\langle \hat{H}_{so} \rangle$ matrix

We have marked the non-zero elements.

Using the operator properties derived earlier we get

$$
\begin{aligned}
a &= \langle 1|\, \hat{H}_{so}\, |1\rangle = \left\langle 2,1,\frac{1}{2},1,\frac{1}{2}\right| \hat{H}_{so} \left|2,1,\frac{1}{2},1,\frac{1}{2}\right\rangle_{mm} \\
&= \left\langle 2,1,\frac{1}{2},1,\frac{1}{2}\right| \frac{e^2}{2m^2c^2}\left(\frac{1}{r^3}\right)\left(\hat{L}_z\hat{S}_z + \frac{1}{2}\left(\hat{L}_+\hat{S}_- + \hat{L}_-\hat{S}_+\right)\right)\left|2,1,\frac{1}{2},1,\frac{1}{2}\right\rangle_{mm} \\
&= \frac{e^2}{2m^2c^2}\langle 2,1|\,\frac{1}{r^3}\,|2,1\rangle_{mm}\left\langle 2,1,\frac{1}{2},1,\frac{1}{2}\right|\left(\hat{L}_z\hat{S}_z + \frac{1}{2}\left(\hat{L}_+\hat{S}_- + \hat{L}_-\hat{S}_+\right)\right)\left|2,1,\frac{1}{2},1,\frac{1}{2}\right\rangle_{mm} \\
&= \frac{e^2}{2m^2c^2}\langle 2,1|\,\frac{1}{r^3}\,|2,1\rangle_{mm}\left\langle 2,1,\frac{1}{2},1,\frac{1}{2}\right| \hat{L}_z\hat{S}_z \left|2,1,\frac{1}{2},1,\frac{1}{2}\right\rangle \\
&= \frac{e^2\hbar^2}{4m^2c^2}\int_0^\infty \frac{1}{r}R_{n\ell}^2(r)dr = \frac{e^2\hbar^2}{4m^2c^2}\left[\frac{Z^3}{a_0^3 n^3 \ell(\ell+\frac{1}{2})(\ell+1)}\right]
\end{aligned}
$$

Similar calculations give $\langle\ \hat{H}_{so}\ \rangle$ as shown in the Table below.

	1	2	3	4	5	6	7	8
1	a	0	0	0	0	0	0	0
2	0	-a	$\sqrt{2}$a	0	0	0	0	0
3	0	$\sqrt{2}$a	0	0	0	0	0	0
4	0	0	0	0	$\sqrt{2}$a	0	0	0
5	0	0	0	$\sqrt{2}$a	-a	0	0	0
6	0	0	0	0	0	a	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0

$\langle \hat{H}_{so} \rangle$ matrix - revised

This says that (only diagonal elements)

$$E_1^{(1)} = a = E_6^{(1)} \quad , \quad E_7^{(1)} = 0 = E_8^{(1)}$$

and

$$|1\rangle = \left|2, 1, \frac{1}{2}, \frac{3}{2}, \frac{3}{2}\right\rangle_{jm} \text{ and } |6\rangle = \left|2, 1, \frac{1}{2}, \frac{3}{2}, -\frac{3}{2}\right\rangle_{jm}$$

$$|7\rangle = \left|2, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm} \text{ and } |8\rangle = \left|2, 0, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right\rangle_{jm}$$

In order to find $E^{(1)}{}_{2'}$, $E^{(1)}{}_{3'}$, $E^{(1)}{}_{4'}$, $E^{(1)}{}_{5'}$, corresponding to the new zero-order state vectors $|2'\rangle$,$|3'\rangle$,$|4'\rangle$,$|5'\rangle$, we must diagonalize the two 2×2 submatrices.

We begin with the submatrix involving states $|2\rangle$ and $|3\rangle$ as shown in the Table, namely,

	2	3
2	-a	$\sqrt{2}$a
3	$\sqrt{2}$a	0

$\langle \hat{H}_{so} \rangle$ 2-3 submatrix

The characteristic equation is

$$(-a - E)(-E) - 2a^2 = 0 = E^2 + aE - 2a^2$$

or

$$E_{2'}^{(1)} = a \text{ and } E_{3'}^{(1)} = -2a$$

Notice that these energies are

$$E_{2'}^{(1)} = \ell a \text{ and } E_{3'}^{(1)} = -(\ell + 1)a$$

as expected for the

$$j = \ell + \frac{1}{2} \text{ and } j = \ell - \frac{1}{2}$$

states respectively.

We find the eigenvectors using the eigenvalue equations.

For $|2'\rangle$ we have

$$\begin{pmatrix} -a & a\sqrt{2} \\ a\sqrt{2} & 0 \end{pmatrix} |2'\rangle = \begin{pmatrix} -a & a\sqrt{2} \\ a\sqrt{2} & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = E^{(1)}_{2'} \begin{pmatrix} u \\ v \end{pmatrix} = a \begin{pmatrix} u \\ v \end{pmatrix}$$

or

$$-u + \sqrt{2}v = u \text{ and } \sqrt{2}u = v$$

Using the normalization condition $u^2 + v^2 = 1$ we get

$$u = \sqrt{\frac{1}{3}} \text{ and } v = \sqrt{\frac{2}{3}}$$

or

$$|2'\rangle = \sqrt{\frac{1}{3}}|2\rangle + \sqrt{\frac{2}{3}}|3\rangle = \left|2, 1, \frac{1}{2}, \frac{3}{2}, \frac{1}{2}\right\rangle_{jm}$$

and similarly

$$|3'\rangle = \sqrt{\frac{2}{3}}|2\rangle - \sqrt{\frac{1}{3}}|3\rangle = \left|2, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm}$$

We then deal with the submatrix involving state $|4\rangle$ and $|5\rangle$ in the Table, namely

	4	5
4	0	$\sqrt{2}$a
5	$\sqrt{2}$a	-a

$\langle \hat{H}_{so} \rangle$ 4-5 submatrix

The characteristic equation is

$$(-a-E)(-E)-2a^2=0=E^2+aE-2a^2$$

or

$$E_{4'}^{(1)}=a \text{ and } E_{5'}^{(1)}=-2a$$

and the eigenvectors are

$$|4'\rangle=\sqrt{\frac{2}{3}}|4\rangle+\sqrt{\frac{1}{3}}|5\rangle=\left|2,1,\frac{1}{2},\frac{3}{2},-\frac{1}{2}\right\rangle_{jm}$$

$$|5'\rangle=\sqrt{\frac{1}{3}}|4\rangle-\sqrt{\frac{2}{3}}|5\rangle=\left|2,1,\frac{1}{2},\frac{1}{2},-\frac{1}{2}\right\rangle_{jm}$$

So including the spin-orbit correction we end up with the energy levels

$$E_2^{(0)}+a \text{ for } 1,2',4',6 \rightarrow \text{ 4 - fold degenerate}$$

$$E_2^{(0)} \text{ for } 7,8 \rightarrow \text{ 2 - fold degenerate}$$

$$E_2^{(0)}-2a \text{ for } 3',5' \rightarrow \text{ 2 - fold degenerate}$$

Spin-Orbit and Arbitrary Magnetic Field

Now let us add on the Zeeman correction (for $\vec{B}=B\,\hat{z}$)

$$\hat{H}_{Zeeman}=\frac{\mu_B}{\hbar}(\vec{L}_{op}+2\vec{S}_{op})\cdot\vec{B}=\frac{\mu_B B}{\hbar}(\hat{L}_z+2S_z)$$

We can solve this problem for arbitrary magnetic field by repeating Method #2 using the correction term as the sum of spin-orbit and Zeeman effects.

The zero-order Hamiltonian is $\hat{H}_0$ and the zero-order state vectors are the $|n, \ell, s, m_\ell, m_s\rangle$ states.

The eight zero-order $n = 2$ states are all degenerate with energy

$$E_2^{(0)} = -\frac{e^2}{8a_0}$$

so we must use degenerate perturbation theory.

We have already calculated the $\langle\ \hat{H}_{so}\ \rangle$ in this basis.

It is shown in the Table below

	1	2	3	4	5	6	7	8
1	a	0	0	0	0	0	0	0
2	0	-a	$\sqrt{2}$a	0	0	0	0	0
3	0	$\sqrt{2}$a	0	0	0	0	0	0
4	0	0	0	0	$\sqrt{2}$a	0	0	0
5	0	0	0	$\sqrt{2}$a	-a	0	0	0
6	0	0	0	0	0	a	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0

$\langle \hat{H}_{so} \rangle$ matrix

where

$$a = \frac{e^2\hbar^2}{96m^2a_0^3c^2}$$

The $\langle\ \hat{L}_z + 2\hat{S}_z\rangle$ matrix is diagonal in this representation and its diagonal elements are given by m_ℓ + $2m_s$ and so the Zeeman contribution $\langle\ \hat{H}_{Zeeman}\rangle$ is shown in the Table below.

	1	2	3	4	5	6	7	8
1	2b	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	b	0	0	0	0	0
4	0	0	0	-b	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	-2b	0	0
7	0	0	0	0	0	0	b	0
8	0	0	0	0	0	0	0	-b

$\langle \hat{H}_{Zeeman} \rangle$ matrix

where $b = \mu_B B$.

The combined perturbation matrix $\langle \quad \hat{V} \quad \rangle$ is then given in the Table below.

	1	2	3	4	5	6	7	8
1	a+2b	0	0	0	0	0	0	0
2	0	-a	$a\sqrt{2}$	0	0	0	0	0
3	0	$a\sqrt{2}$	b	0	0	0	0	0
4	0	0	0	-b	$a\sqrt{2}$	0	0	0
5	0	0	0	$a\sqrt{2}$	-a	0	0	0
6	0	0	0	0	0	a-2b	0	0
7	0	0	0	0	0	0	b	0
8	0	0	0	0	0	0	0	-b

$\langle \hat{V} \rangle$ matrix

After diagonalizing, the new energies are

$$E_{1'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} + 2\mu_B B$$

$$E_{2'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) + \sqrt{9a^2 + 2ab + b^2})$$

$$E_{3'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) - \sqrt{9a^2 + 2ab + b^2})$$

$$E_{4'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) + \sqrt{9a^2 - 2ab + b^2})$$

$$E_{5'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) - \sqrt{9a^2 - 2ab + b^2})$$

$$E_{6'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} - 2\mu_B B$$

$$E_{7'} = -\frac{e^2}{8a_0} + \mu_B B$$

$$E_{8'} = -\frac{e^2}{8a_0} - \mu_B B$$

If we let B be small so that $b \ll a$, we then get the approximate energies

$$E_{1'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} + 2\mu_B B$$

$$E_{2'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) + 3a(1+\frac{b}{9a}))$$

$$= -\frac{e^2}{8a_0} + a + \frac{2}{3}b = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} + \frac{2}{3}\mu_B B$$

$$E_{3'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) - 3a(1+\frac{b}{9a}))$$

$$= -\frac{e^2}{8a_0} - 2a + \frac{1}{3}b = -\frac{e^2}{8a_0} - \frac{e^2}{8a_0}\frac{\alpha^2}{6} + \frac{1}{3}\mu_B B$$

$$E_{4'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) + 3a(1-\frac{b}{9a}))$$

$$= -\frac{e^2}{8a_0} + a - \frac{2}{3}b = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} - \frac{2}{3}\mu_B B$$

$$E_{5'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) - 3a(1 - \frac{b}{9a}))$$

$$= -\frac{e^2}{8a_0} - 2a - \frac{1}{3}b = -\frac{e^2}{8a_0} - \frac{e^2}{8a_0}\frac{\alpha^2}{6} - \frac{1}{3}\mu_B B$$

$$E_{6'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} - 2\mu_B B$$

$$E_{7'} = -\frac{e^2}{8a_0} + \mu_B B$$

$$E_{8'} = -\frac{e^2}{8a_0} - \mu_B B$$

This is clearly a perturbation of the spin-orbit energy levels.

We assume that the new state vectors become the zero-order vectors in the spin-orbit case for low fields.

In this case the Zeeman effect corrections(to the fine structure energies) are given by the Table below.

ℓ	s	j	m_j	ΔE_{Zeeman}	State
1	1/2	3/2	3/2	$2\mu_B B$	$1'$
1	1/2	3/2	1/2	$2\mu_B B/3$	$2'$
1	1/2	1/2	1/2	$\mu_B B/3$	$3'$
1	1/2	3/2	-1/2	$-2\mu_B B/3$	$4'$
1	1/2	1/2	-1/2	$-\mu_B B/3$	$5'$
1	1/2	3/2	-3/2	$-2\mu_B B$	$6'$
0	1/2	1/2	1/2	$\mu_B B$	$7'$
0	1/2	1/2	-1/2	$-\mu_B B$	$8'$

n = 2 energy corrections for small B

A little bit of study shows the general relation

$$\Delta E_{Zeeman} = g\mu_B B m_j$$

where

$$g = \text{ Lande g - factor } = 1 + \frac{j(j+1) - \ell(\ell+1) + s(s+1)}{2j(j+1)}$$

This is called the *Zeeman effect*.

We can prove this result in general.

The general method uses the *Wigner-Eckart Theorem*.

Wigner-Eckart Theorem

Consider a vector operator $\vec{A}_{op}$.

We have already shown that the Cartesian components of any vector operator has the following commutation relations with the Cartesian components of the angular momentum operator

$$\left[\hat{A}_i, \hat{J}_j\right] = i\hbar\varepsilon_{ijk}\hat{A}_k$$

We will now prove the following powerful theorem:

In a basis that diagonalizes $\vec{J}^2_{op}$ and $\hat{J}_z$ (i.e., the $|\lambda, \ell, s, j, m_j\rangle$ states, where λ signifies other operators that commute with $\vec{J}^2_{op}$ and $\hat{J}_z$), the matrix elements of $\vec{A}_{op}$ between states with the same j–value are proportional to the matrix elements of $\vec{J}_{op}$ and the proportionality factor is independent of m_j.

The algebra involved in the proof is simpler if we work in the so-called *spherical basis* instead of the Cartesian basis.

The spherical basis uses

$$\hat{J}_\pm = \hat{J}_x \pm i\hat{J}_y \quad , \quad \hat{J}_0 = \hat{J}_z$$
$$\hat{A}_\pm = \hat{A}_x \pm i\hat{A}_y \quad , \quad \hat{A}_0 = \hat{A}_z$$

The corresponding commutators are

$$\left[\hat{A}_\pm, \hat{J}_0\right] = \mp\hbar\hat{A}_\pm \quad , \quad \left[\hat{A}_\pm, \hat{J}_\pm\right] = 0$$
$$\left[\hat{A}_\pm, \hat{J}_\mp\right] = \pm 2\hbar\hat{A}_0 \quad , \quad \left[\hat{A}_0, \hat{J}_0\right] = 0$$
$$\left[\hat{A}_0, \hat{J}_\pm\right] = \pm\hbar\hat{A}_\pm$$

which all follow from the original commutator for the arbitrary vector operator.

Now, by definition of the operators, we have

$$\hat{J}_0 |\lambda, j, m_j\rangle = \hbar m_j |\lambda, j, m_j\rangle = \langle j, m_j| \hat{J}_0 |j, m_j\rangle |\lambda, j, m_j\rangle$$
$$\hat{J}_\pm |\lambda, j, m_j\rangle = \hbar\sqrt{j(j+1) - m_j(m_j \pm 1)} |\lambda, j, m_j \pm 1\rangle$$
$$= \langle j, m_j \pm 1| \hat{J}_\pm |j, m_j\rangle |\lambda, j, m_j \pm 1\rangle$$
$$\langle \lambda, j, m_j| \hat{J}_\pm = \left(\hat{J}_\mp |\lambda, j, m_j\rangle\right)^+$$
$$= \langle \lambda, j, m_j \mp 1| \hbar\sqrt{j(j+1) - m_j(m_j \mp 1)}$$
$$= \langle \lambda, j, m_j \mp 1| \langle j, m_j| \hat{J}_\pm |j, m_j \mp 1\rangle$$

We now work with the matrix elements of some of the commutators and use the defining relations above to prove the theorem.

First, we have

$$\mp\hbar\left\langle\lambda',j,m_j'\right|\hat{A}_\pm\left|\lambda,j,m_j\right\rangle=\left\langle\lambda',j,m_j'\right|\left[\hat{A}_\pm,\hat{J}_0\right]\left|\lambda,j,m_j\right\rangle$$
$$=\left\langle\lambda',j,m_j'\right|\left[\hat{A}_\pm\hat{J}_0-\hat{J}_0\hat{A}_\pm\right]\left|\lambda,j,m_j\right\rangle$$
$$=(m_j-m_j')\hbar\left\langle\lambda',j,m_j'\right|\hat{A}_\pm\left|\lambda,j,m_j\right\rangle$$

or

$$0=(m_j-m_j'\pm1)\hbar\left\langle\lambda',j,m_j'\right|\hat{A}_\pm\left|\lambda,j,m_j\right\rangle$$

This says that either $m'_{\mathrm{j}}=m_{\mathrm{j}}\pm1$ or the matrix element

$$\left\langle\lambda',j,m_j'\right|\hat{A}_\pm\left|\lambda,j,m_j\right\rangle=0$$

Since we have an identical property for the matrix elements of $\hat{J}_\pm$ this implies that the matrix elements of $\hat{A}_\pm$ are proportional to those of $\hat{J}_\pm$ and we can write the proportionality constant as

$$\frac{\left\langle\lambda',j,m_j\pm1\right|\hat{A}_\pm\left|\lambda,j,m_j\right\rangle}{\left\langle j,m_j\pm1\right|\hat{J}_\pm\left|j,m_j\right\rangle}$$

Second, we have

$$\left\langle\lambda',j,m_j'\right|\left[\hat{A}_\pm,\hat{J}_\pm\right]\left|\lambda,j,m_j\right\rangle=0$$
$$\left\langle\lambda',j,m_j'\right|\hat{A}_\pm\hat{J}_\pm\left|\lambda,j,m_j\right\rangle=\left\langle\lambda',j,m_j'\right|\hat{J}_\pm\hat{A}_\pm\left|\lambda,j,m_j\right\rangle$$
$$\left\langle\lambda',j,m_j'\right|\hat{A}_\pm\left|\lambda,j,m_j\pm1\right\rangle\left\langle\lambda',j,m_j\pm1\right|\hat{J}_\pm\left|\lambda,j,m_j\right\rangle$$
$$=\left\langle\lambda',j,m_j'\mp1\right|\hat{A}_\pm\left|\lambda,j,m_j\right\rangle\left\langle\lambda,j,m_j\pm1\right|\hat{J}_\pm\left|\lambda,j,m_j'\mp1\right\rangle$$

Using the result from the first commutator this says that $m'_{\mathrm{j}}=m_{\mathrm{j}}\pm2$, which, in turn, implies that

$$\frac{\left\langle\lambda',j,m_j\pm2\right|\hat{A}_\pm\left|\lambda,j,m_j\pm1\right\rangle}{\left\langle j,m_j\pm2\right|\hat{J}_\pm\left|j,m_j\pm1\right\rangle}=\frac{\left\langle\lambda',j,m_j\pm1\right|\hat{A}_\pm\left|\lambda,j,m_j\right\rangle}{\left\langle j,m_j\pm1\right|\hat{J}_\pm\left|j,m_j\right\rangle}$$

This says that the proportionality constant is independent of m_j.

We define a new symbol for the proportionality constant

$$\langle \lambda', j || A || \lambda, j \rangle_{\pm} = \text{ the reduced matrix element}$$

which gives the relation

$$\langle \lambda', j, m_j' | \hat{A}_{\pm} | \lambda, j, m_j \rangle = \langle \lambda', j || A || \lambda, j \rangle_{\pm} \langle j, m_j' | \hat{J}_{\pm} | j, m_j \rangle$$

To complete the proof we need to show that the same result holds for $\hat{A}_0$ and that

$$\langle \lambda', j || A || \lambda, j \rangle_{+} = \langle \lambda', j || A || \lambda, j \rangle_{-}$$

We have

$$\begin{aligned} \pm 2\hbar \langle \lambda', j, m_j' | \hat{A}_0 | \lambda, j, m_j \rangle &= \langle \lambda', j, m_j' | [\hat{A}_{\pm}, \hat{J}_{\mp}] | \lambda, j, m_j \rangle \\ &= \langle \lambda', j, m_j' | [\hat{A}_{\pm} \hat{J}_{\mp} - \hat{J}_{\mp} \hat{A}_{\pm}] | \lambda, j, m_j \rangle \\ &= \langle \lambda', j, m_j' | \hat{A}_{\pm} | \lambda, j, m_j \mp 1 \rangle \langle j, m_j \mp 1 | \hat{J}_{\mp} | j, m_j \rangle \\ &\quad - \langle \lambda', j, m_j' \pm 1 | \hat{A}_{\pm} | \lambda, j, m_j \rangle \langle j, m_j' | \hat{J}_{\mp} | j, m_j' \pm 1 \rangle \end{aligned}$$

Now substituting in the matrix element of $\hat{A}_{\pm}$ we get

$$\begin{aligned} &\pm 2\hbar \langle \lambda', j, m_j' | \hat{A}_0 | \lambda, j, m_j \rangle \\ &\quad = \langle \lambda', j || A || \lambda, j \rangle_{\pm} [\langle j, m_j' | \hat{J}_{\pm} | j, m_j \mp 1 \rangle \langle j, m_j \mp 1 | \hat{J}_{\mp} | j, m_j \rangle \\ &\qquad - \langle j, m_j' \pm 1 | \hat{J}_{\pm} | j, m_j \rangle \langle j, m_j' | \hat{J}_{\mp} | j, m_j' \pm 1 \rangle] \end{aligned}$$

This says that $\hat{A}_0$ has non-vanishing matrix elements only when $m'_j = m_j$. We then get

$$\pm 2\hbar \left\langle \lambda', j, m_j' \right| \hat{A}_0 \left| \lambda, j, m_j \right\rangle$$

$$= \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle_{\pm} \left[\left| \left\langle j, m_j \mp 1 \right| \hat{J}_{\mp} \left| j, m_j \right\rangle \right|^2 - \left| \left\langle j, m_j \pm 1 \right| \hat{J}_{\mp} \left| j, m_j \right\rangle \right|^2 \right]$$

$$= \pm 2\hbar m_j = \pm 2\hbar \left\langle \lambda', j, m_j' \right| \hat{J}_0 \left| \lambda, j, m_j \right\rangle$$

Putting it all together we get

$$\left\langle \lambda', j, m_j' \right| \hat{A}_0 \left| \lambda, j, m_j \right\rangle = \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle_{\pm} \left\langle j, m_j' \right| \hat{J}_0 \left| j, m_j \right\rangle$$

Since no operator has a ± subscript, this also says that

$$\left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle_{+} = \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle_{-} = \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle$$

and we finally have

$$\left\langle \lambda', j, m_j' \right| \vec{A}_{op} \left| \lambda, j, m_j \right\rangle = \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle_{\pm} \left\langle j, m_j' \right| \vec{J}_{op} \left| j, m_j \right\rangle$$

This completes the proof of the *Wigner-Eckart theorem*.

A very important extension of this theorem is the following result:

$$\left\langle \lambda', j, m_j' \right| \vec{A}_{op} \cdot \vec{J}_{op} \left| \lambda, j, m_j \right\rangle = \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \left\langle j, m_j' \right| \vec{J}_{op}^2 \left| j, m_j \right\rangle$$

$$= \delta_{m_j', m_j} \hbar^2 j(j+1) \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle$$

which says that the scalar product is diagonal in m_{j} .

This result follows directly from the Wigner-Eckart theorem

$$\left\langle \lambda', j, m_j' \right| \vec{A}_{op} \cdot \vec{J}_{op} \left| \lambda, j, m_j \right\rangle = \sum_k \left\langle \lambda', j, m_j' \right| A_{op,k} J_{op,k} \left| \lambda, j, m_j \right\rangle$$

$$= \sum_k \left\langle \lambda', j, m_j' \right| A_{op,k} \left(\sum_{m_j''} \left| \lambda, j, m_j'' \right\rangle \left\langle \lambda, j, m_j'' \right| \right) J_{op,k} \left| \lambda, j, m_j \right\rangle$$

$$= \sum_{\lambda'', j'', m_j''} \sum_k \left\langle \lambda', j, m_j' \right| A_{op,k} \left| \lambda, j, m_j'' \right\rangle \left\langle \lambda, j, m_j'' \right| J_{op,k} \left| \lambda, j, m_j \right\rangle$$

$$= \sum_{m_j''} \sum_k \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \left\langle \lambda', j, m_j' \right| J_{op,k} \left| \lambda, j, m_j'' \right\rangle \left\langle \lambda, j, m_j'' \right| J_{op,k} \left| \lambda, j, m_j \right\rangle$$

$$= \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \sum_k \left\langle \lambda', j, m_j' \right| J_{op,k} \left(\sum_{m_j''} \left| \lambda, j, m_j'' \right\rangle \left\langle \lambda, j, m_j'' \right| \right) J_{op,k} \left| \lambda, j, m_j \right\rangle$$

$$= \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \left\langle j, m_j' \right| \vec{J}_{op}^2 \left| j, m_j \right\rangle = \delta_{m_j', m_j} \hbar^2 j(j+1) \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle$$

Now back to the Zeeman effect. In the low field limit, we need to evaluate the diagonal matrix elements

$$\left\langle \ell s j m_j \right| (\hat{L}_z + 2\hat{S}_z) \left| \ell s j m_j \right\rangle = \left\langle \ell s j m_j \right| (\hat{J}_z + \hat{S}_z) \left| \ell s j m_j \right\rangle$$

$$= \hbar m_j + \left\langle \ell s j m_j \right| \hat{S}_z \left| \ell s j m_j \right\rangle$$

Now the Wigner-Eckart theorem says that

$$\left\langle \ell s j m_j \right| \hat{S}_z \left| \ell s j m_j \right\rangle = \left\langle \ell s j \right| |S| \left| \ell s j \right\rangle \left\langle \ell s j m_j \right| \hat{J}_z \left| \ell s j m_j \right\rangle$$

$$= \hbar m_j \left\langle \ell s j \right| |S| \left| \ell s j \right\rangle$$

The scalar product matrix element formula gives

$$\langle \ell s j m_j | \vec{S}_{op} \cdot \vec{J}_{op} | \ell s j m_j \rangle = \langle \ell s j | | S | | \ell s j \rangle \langle j m_j | \vec{J}_{op}^2 | j m_j \rangle$$
$$= \hbar^2 j(j+1) \langle \ell s j | | S | | \ell s j \rangle$$

But we also have

$$(\vec{J}_{op} - \vec{S}_{op})^2 = \vec{L}_{op}^2 = \vec{J}_{op}^2 + \vec{S}_{op}^2 - 2\vec{S}_{op} \cdot \vec{J}_{op}$$

$$\langle \ell s j m_j | \vec{S}_{op} \cdot \vec{J}_{op} | \ell s j m_j \rangle = \frac{1}{2} \langle \ell s j m_j | (\vec{J}_{op}^2 + \vec{S}_{op}^2 - \vec{L}_{op}^2) | \ell s j m_j \rangle$$
$$= \frac{1}{2} \hbar^2 (j(j+1) + s(s+1) - \ell(\ell+1))$$

or

$$\langle \ell s j | | S | | \ell s j \rangle = \frac{j(j+1) + s(s+1) - \ell(\ell+1)}{2j(j+1)}$$

and thus

$$\langle \ell s j m_j | (\hat{L}_z + 2\hat{S}_z) | \ell s j m_j \rangle$$
$$= \hbar m_j + \hbar m_j \frac{j(j+1) + s(s+1) - \ell(\ell+1)}{2j(j+1)}$$
$$= \hbar m_j g_{j\ell s}$$

$$g_{j\ell s} = 1 + \frac{j(j+1) + s(s+1) - \ell(\ell+1)}{2j(j+1)} = \text{ Lande g - factor}$$

Finally, we have

$$\langle \ell s j m_j | \hat{H}_{Zeeman} | \ell s j m_j \rangle = \mu_B B m_j g_{j\ell s}$$

and the result we found earlier in the special example case is now proved in general.

Paschen-Bach Effect

When B is large enough such that $\Delta E_{\text{Zeeman}} \gg \Delta E_{\text{so}}$, but not large enough so that the $\vec{B}^2$ term we neglected earlier is important. We have the so-called *Paschen-Bach effect*.

If the $\vec{B}^2$ term is dominant we have the so-called *quadratic Zeeman effect*.

The best way to see what is happening for all magnetic field values is a plot.

In CGS Gaussian units $\mu_B = 5.7884 \times 10^{-9} \dfrac{eV}{gauss}$, $a_0 = 5.2918 \times 10^{-8}\, cm$, $e = 4.80 \times 10^{-10}\, esu$

$$\frac{e^2}{a_0} = 27.2\, eV a = 1.509 \times 10^{-5}\, eV b = 5.7884 \times 10^{-9} B\, eV$$

Using our earlier results we then have

$$E_{1'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} + 2\mu_B B \qquad E_{5'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) - \sqrt{9a^2 - 2ab + b^2})$$

$$E_{2'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) + \sqrt{9a^2 + 2ab + b^2}) \qquad E_{6'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} - 2\mu_B B$$

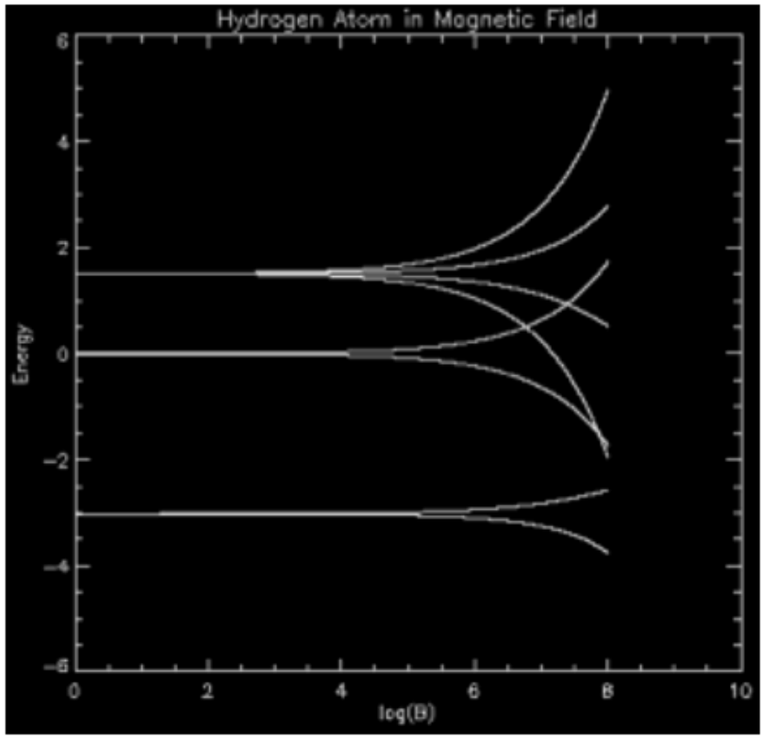
$$E_{3'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) - \sqrt{9a^2 + 2ab + b^2}) \qquad E_{7'} = -\frac{e^2}{8a_0} + \mu_B B$$

$$E_{4'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) + \sqrt{9a^2 - 2ab + b^2}) \qquad E_{8'} = -\frac{e^2}{8a_0} - \mu_B B$$

A plot of

$$\left(E + \frac{e^2}{8a_0}\right) \times 10^5\, eV \text{ versus } \log_e(B(gauss))$$

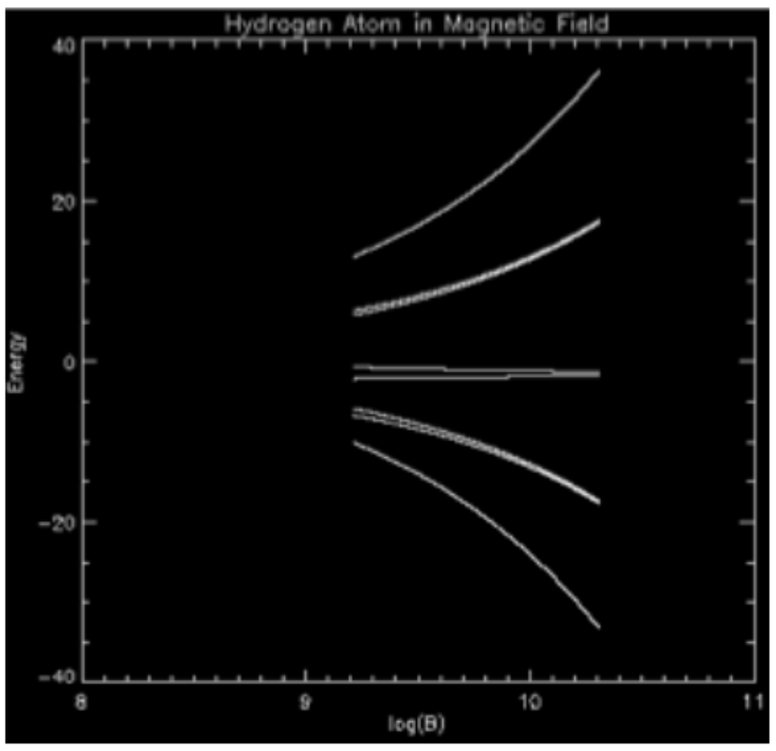
looks like the Figure below.



Hydrogen Atom In a Magnetic Field - Zeeman Effect

This plot for fields below 400 gauss ($\log_e(B) \approx 6$) shows the characteristic level structure of the Zeeman effect.

The very large magnetic field Paschen-Bach effect is illustrated in the Figure below.



Hydrogen Atom In a Magnetic Field - Paschen-Bach Effect

Notice the equally-spaced level signature of the Paschen-Bach effect.

We now define some notation that will be important in the study atomic spectra.

For small magnetic fields we found that the approximate state vectors are the $|n\ell sjm_j\rangle$ states.

The energy levels including spin-orbit effects are

$$\begin{aligned} E_n &= E_n^{(0)} + \Delta E_{so} \\ &= -\frac{Ze^2}{2a_0 n^2} \\ &\quad + \frac{Z^2\alpha^2 \left|E_n^{(0)}\right|}{n\ell(2\ell+1)(\ell+1)}(1-\delta_{\ell,0})\left(\ell\delta_{j,\ell+1/2} - (\ell+1)\delta_{j,\ell-1/2}\right) \end{aligned}$$

We define a spectroscopic notation to label the energy levels using the scheme shown below

$$|n\ell sjm_j\rangle \rightarrow n^{2S+1}L(symbol)_J$$

so that

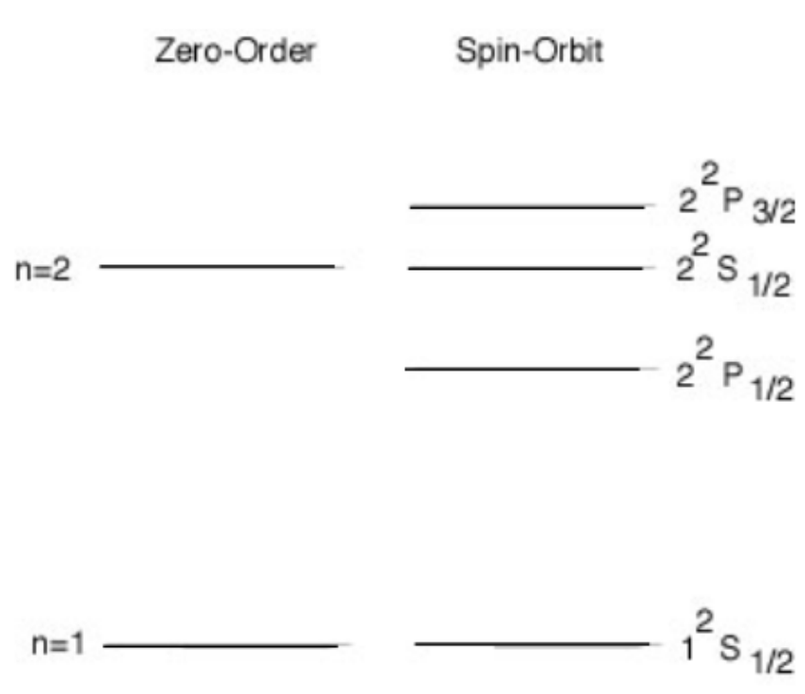
$$\left|21\frac{1}{2}\frac{3}{2}m_j\right\rangle \rightarrow 2^2P_{\frac{3}{2}}\ ,\ \left|21\frac{1}{2}\frac{1}{2}m_j\right\rangle \rightarrow 2^2P_{\frac{1}{2}}\ ,\ \left|20\frac{1}{2}\frac{1}{2}m_j\right\rangle \rightarrow 2^2S_{\frac{1}{2}}$$

The *L*(symbols) are defined by the Table.

L	0	1	2	3
Symbol	S	P	D	F

Spectroscopic Labels

The energy level diagram for $n = 1$ and $n = 2$ is shown in the Figure below.



Spin-Orbit Energy Levels

Earlier we calculated the relativistic correction and found that it was the same order of magnitude as the spin-orbit correction for hydrogen.

We found

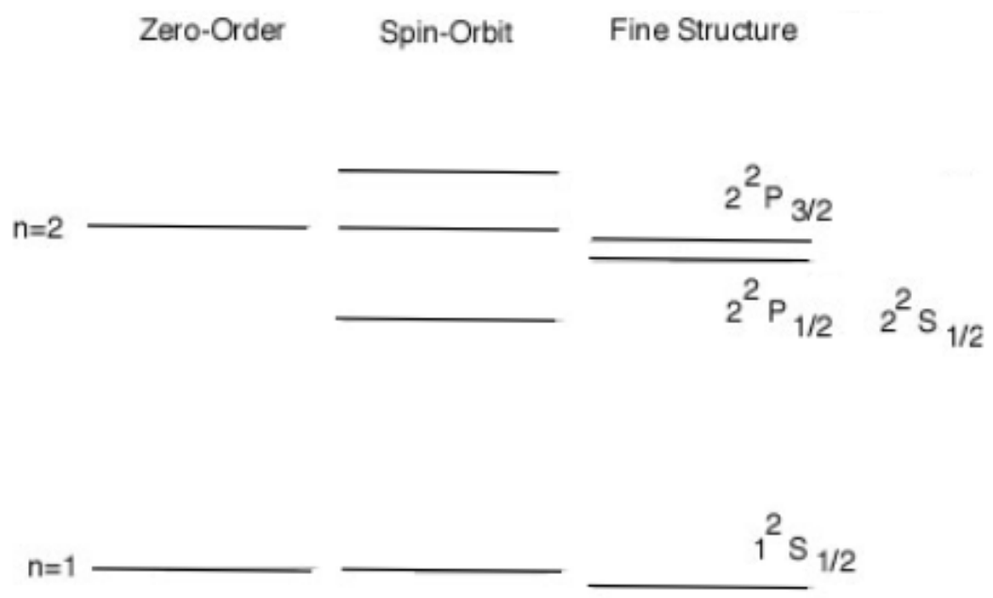
$$\Delta E_{rel} = -\frac{Z^2\alpha^2\left|E_n^{(0)}\right|}{n}\left(\frac{2}{2\ell+1}-\frac{3}{4n}\right)$$

Combining these two corrections we have

$$\Delta E_{fs} = \Delta E_{so} + \Delta E_{rel} = -\frac{Z^2\alpha^2\left|E_n^{(0)}\right|}{n}\left(\frac{1}{j+\frac{1}{2}}-\frac{3}{4n}\right) \quad j = \ell \pm \frac{1}{2}$$

which is independent of ℓ .

This changes the energy level structure to that shown in the Figure below.



Fine Structure Energy Levels

The observed spectral lines result from an electron making a transition between these levels.

Stark Effect

When a hydrogen atom is placed in an external electric field $\vec{\mathcal{E}}$, the potential energy of the proton and the electron is given by

$$V_{dipole}(\vec{r}_e, \vec{r}_p) = -e\vec{\mathcal{E}}_0 \cdot \vec{r}_p + e\vec{\mathcal{E}}_0 \cdot \vec{r}_e$$
$$= e\mathcal{E}_0(z_e - z_p) = e\mathcal{E}_0 z$$

where

$$z = z_e - z_p = z_{relative}$$

Therefore, we consider the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}_{dipole}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{op}^2}{2m} - e^2\left(\frac{1}{r}\right)_{op}$$

$$\hat{H}_{dipole} = e\mathcal{E}_0 z_{op}$$

For weak electric fields, we can apply perturbation theory (we ignore spin in this calculation).

First, we apply perturbation theory to the $n = 1$ ground state of the hydrogen atom.

For the ground state, the wave function is $\psi_{100}(\vec{r})$ and the first-order correction to the energy is

$$\begin{aligned}
E_1^{(1)} &= \langle 100 | \, e\mathcal{E}_0 z_{op} \, | 100 \rangle \\
&= e\mathcal{E}_0 \int d^3\vec{r} d^3\vec{r}' \, \langle 100 \mid \vec{r} \rangle \, \langle \vec{r} | \, z_{op} \, | \vec{r}' \rangle \, \langle \vec{r}' \mid 100 \rangle \\
&= e\mathcal{E}_0 \int d^3\vec{r} d^3\vec{r}' z \psi_{100}^*(\vec{r}) \, \langle \vec{r} \mid \vec{r}' \rangle \, \psi_{100}(\vec{r}') \\
&= e\mathcal{E}_0 \int d^3\vec{r} d^3\vec{r}' z \psi_{100}^*(\vec{r}) \delta(\vec{r} - \vec{r}') \psi_{100}(\vec{r}') \\
&= e\mathcal{E}_0 \int d^3\vec{r} z \, |\psi_{100}(\vec{r})|^2
\end{aligned}$$

This equals zero since the integrand is the product of an even and odd functions.

Thus, the first-order correction is zero for the ground state.

The second-order correction is given by non-degenerate perturbation theory as

$$E_1^{(2)} = \sum_{n=2}^{\infty} \sum_{\ell=0}^{n-1} \sum_{m=-\ell}^{\ell} \frac{|\langle n\ell m | \, e\mathcal{E}_0 z_{op} \, | 100 \rangle|^2}{E_1^{(0)} - E_m^{(0)}}$$

Using $z = r \cos\theta$ we have

$$\langle n\ell m | \, z_{op} \, | 100 \rangle = \int d^3\vec{r} \, [R_{n\ell}(r) Y_{\ell m}^*(\theta, \phi)] \, [r \cos\theta] \, R_{10}(r) Y_{00}(\theta, \phi)$$

Now

$$Y_{00} = \frac{1}{\sqrt{4\pi}} \text{ and } z = \sqrt{\frac{4\pi}{3}} Y_{10}$$

Therefore

$$\langle n\ell m|\, z_{op}\,|100\rangle = \int r^3 dr R_{n\ell}(r) R_{10}(r) \frac{1}{\sqrt{3}} \int d\Omega Y^*_{\ell m}(\theta,\phi) Y_{10}(\theta,\phi)$$

Now

$$\int d\Omega Y^*_{\ell m}(\theta,\phi) Y_{10}(\theta,\phi) = \delta_{\ell,1}\delta_{m,0}$$

by the orthonormality of the ($\vec{L}^2_{\text{op}}$, $\hat{L}_z$) eigenfunctions.

Therefore,

$$\langle n\ell m|\, z_{op}\,|100\rangle = \frac{1}{\sqrt{3}}\delta_{\ell,1}\delta_{m,0} \int_0^\infty r^3 dr R_{n1}(r) R_{10}(r)$$

and

$$\left|\langle n10|\, z_{op}\,|100\rangle\right|^2 = \frac{1}{3}\frac{2^8 n^7 (n-1)^{2n-5}}{(n+1)^{2n+5}} a_0^2 = \beta(n) a_0^2$$

Finally

$$E_1^{(2)} = (e\mathcal{E}_0 a_0)^2 \sum_{n=2}^{\infty} \frac{\beta(n)}{\frac{e^2}{2a_0}\left(1-\frac{1}{n^2}\right)} = -2F\mathcal{E}_0^2 a_0^3$$

where

$$F = \sum_{n=2}^{\infty} \frac{n^2\beta(n)}{(n^2-1)} \approx 1.125$$

Therefore, the ground state exhibits a *quadratic* Stark effect.

The $n = 2$ level, which is the first excited state of hydrogen, has 4 degenerate states.

$$n = 2 \rightarrow \ell = 0 \rightarrow \psi_{200} = \psi_1$$

$$\rightarrow \ell = 1 \rightarrow m = \begin{cases} 1 \rightarrow \psi_{211} = \psi_2 \\ 0 \rightarrow \psi_{210} = \psi_3 \\ -1 \rightarrow \psi_{21-1} = \psi_4 \end{cases}$$

We must use degenerate perturbation theory.

We construct the 4×4 $\langle\ e\mathcal{E}_0 z_{op}\ \rangle$ matrix and then diagonalize it.

We have

$$\langle e\mathcal{E}_0 z_{op} \rangle = e\mathcal{E}_0 \begin{pmatrix} \langle 1| z_{op} |1\rangle & \langle 1| z_{op} |2\rangle & \langle 1| z_{op} |3\rangle & \langle 1| z_{op} |4\rangle \\ \langle 2| z_{op} |1\rangle & \langle 2| z_{op} |2\rangle & \langle 2| z_{op} |3\rangle & \langle 2| z_{op} |4\rangle \\ \langle 3| z_{op} |1\rangle & \langle 3| z_{op} |2\rangle & \langle 3| z_{op} |3\rangle & \langle 3| z_{op} |4\rangle \\ \langle 4| z_{op} |1\rangle & \langle 4| z_{op} |2\rangle & \langle 4| z_{op} |3\rangle & \langle 4| z_{op} |4\rangle \end{pmatrix}$$

Now z has no ϕ dependence and therefore,

$$\langle j| z_{op} |k\rangle = 0 \text{ if } m_j \neq m_k$$

Thus,

$$\langle 1| z_{op} |2\rangle = 0 = \langle 1| z_{op} |4\rangle$$
$$\langle 2| z_{op} |1\rangle = 0 = \langle 2| z_{op} |3\rangle = \langle 2| z_{op} |4\rangle$$
$$\langle 3| z_{op} |2\rangle = 0 = \langle 3| z_{op} |4\rangle$$
$$\langle 4| z_{op} |1\rangle = 0 = \langle 4| z_{op} |2\rangle = \langle 4| z_{op} |3\rangle$$

and the matrix becomes

$$\langle e\mathcal{E}_0 z_{op} \rangle = e\mathcal{E}_0 \begin{pmatrix} \langle 1| z_{op} |1\rangle & 0 & \langle 1| z_{op} |3\rangle & 0 \\ 0 & \langle 2| z_{op} |2\rangle & 0 & 0 \\ \langle 3| z_{op} |1\rangle & 0 & \langle 3| z_{op} |3\rangle & 0 \\ 0 & 0 & 0 & \langle 4| z_{op} |4\rangle \end{pmatrix}$$

We also have

$$\langle 1| z_{op} |1\rangle = 0 = \langle 2| z_{op} |2\rangle = \langle 3| z_{op} |3\rangle = \langle 4| z_{op} |4\rangle$$

since these integrands involve the product of even and odd functions.

Finding out which matrix elements are equal to zero without actually evaluating the integrals corresponds to finding what are called *selection rules*.

We will elaborate on the idea of selection rules in the next section on the Van der Waal's interaction.

Thus, the matrix finally becomes (after relabeling the rows and columns)

$$\langle e\mathcal{E}_0 z_{op}\rangle = e\mathcal{E}_0 \begin{pmatrix} 0 & \langle 1|\, z_{op}\, |3\rangle & 0 & 0 \\ \langle 3|\, z_{op}\, |1\rangle & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where

$$\langle 1|\, z_{op}\, |3\rangle = \langle 3|\, z_{op}\, |1\rangle = \int \psi_{200}(\vec{r}) z \psi_{210}(\vec{r}) d^3\vec{r} = -3e\mathcal{E}_0 a_0$$

Diagonalizing the 2×2 submatrix gives eigenvalues $\pm 3e\mathcal{E}_0 a_0$.

The first-order energies and new zero-order wave functions are

$$\psi_{211}(\vec{r}) \to E_{211} = E_2^{(0)} \text{ remains degenerate}$$

$$\psi_{21-1}(\vec{r}) \to E_{21-1} = E_2^{(0)} \text{ remains degenerate}$$

$$\psi_{(+)}(\vec{r}) = \frac{1}{\sqrt{2}}\left(\psi_{200}(\vec{r}) - \psi_{210}(\vec{r})\right) \to E_+ = E_2^{(0)} + 3e\mathcal{E}_0 a_0$$

$$\psi_{(-)}(\vec{r}) = \frac{1}{\sqrt{2}}\left(\psi_{200}(\vec{r}) + \psi_{210}(\vec{r})\right) \to E_- = E_2^{(0)} - 3e\mathcal{E}_0 a_0$$

The degeneracy is broken for the $m = 0$ levels and we see a *linear* Stark effect.

The linear Stark effect only appears for degenerate levels.

Van der Waal's Interaction

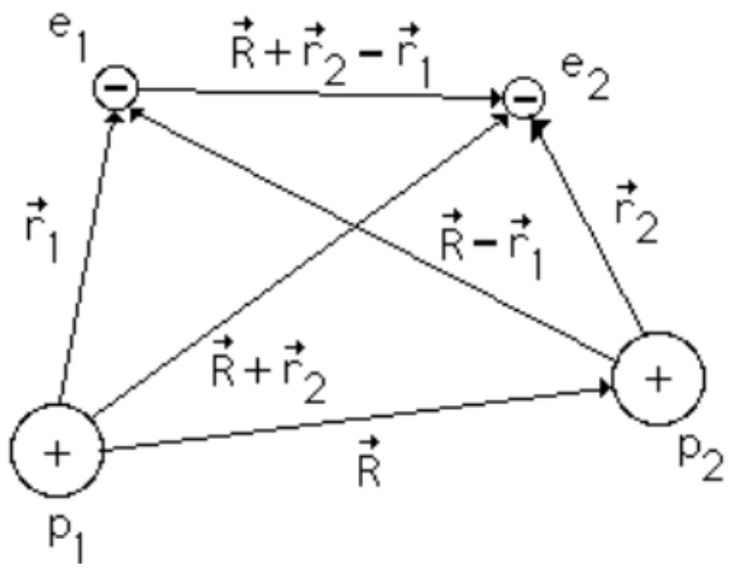
We now consider a system consisting of two widely separated atoms.

In particular, we consider the interaction between two hydrogen atoms, where we treat the two protons as fixed point charges separated by a vector $\vec{R}$ and we define

$\vec{r}_1$ = vector from first proton to its electron

$\vec{r}_2$ = vector from second proton to its electron

as shown in the Figure.



Van der Wall's System

The Hamiltonian is given by

$$\hat{H} = \hat{H}_0 + V$$

where

$$\hat{H}_0 = \frac{\vec{p}_{1,op}^{\,2}}{2\mu} - \frac{e^2}{r_1} + \frac{\vec{p}_{2,op}^{\,2}}{2\mu} - \frac{e^2}{r_2} = \text{ 2 non-interacting hydrogen atoms}$$

and

$$\hat{V} = \text{ rest of the Coulomb interactions}$$

$$= V_{p_1p_2} + V_{e_1e_2} + V_{e_1p_2} + V_{e_2p_1}$$

$$= e^2\left(\frac{1}{R} + \frac{1}{\left|\vec{R} + \vec{r}_2 - \vec{r}_1\right|} - \frac{1}{\left|\vec{R} + \vec{r}_2\right|} - \frac{1}{\left|\vec{R} - \vec{r}_1\right|}\right)$$

This is the perturbation potential.

We know the zero-order solution for two non-interacting hydrogen atoms.

It is

$$\text{zero-order states}: \ |n_1\ell_1 m_1\rangle\,|n_2\ell_2 m_2\rangle$$

with

$$\text{zero-order energies}: \ E^{(0)}_{n_1 n_2} = -\frac{e^2}{2a_0}\left(\frac{1}{n_1^2}+\frac{1}{n_2^2}\right)$$

where

$$\hat{H}_0\,|n_1\ell_1 m_1\rangle\,|n_2\ell_2 m_2\rangle = -\frac{e^2}{2a_0}\left(\frac{1}{n_1^2}+\frac{1}{n_2^2}\right)|n_1\ell_1 m_1\rangle\,|n_2\ell_2 m_2\rangle$$

This expression for the perturbation potential is too complicated to calculate.

We will need to make an approximation.

We make the reasonable assumption that

$$R >> r_2 \text{ and } R >> r_1$$

We have two useful mathematical results that we can apply.

In general, we can write

$$\frac{1}{\left|\vec{R}+\vec{a}\right|} = \frac{1}{\left[R^2+2\vec{R}\cdot\vec{a}+a^2\right]^{1/2}} = \frac{1}{R}\left[1+\frac{2\vec{R}\cdot\vec{a}}{R^2}+\frac{a^2}{R^2}\right]^{-1/2}$$

and for small x we have

$$[1+x]^{-1/2} \approx 1-\frac{1}{2}x+\frac{3}{8}x^2-\frac{5}{16}x^3+\ldots\ldots$$

Using

$$x = \frac{2\vec{R}\cdot\vec{a}}{R^2} + \frac{a^2}{R^2}$$

we get the general result

$$\frac{1}{\left|\vec{R}+\vec{a}\right|} = \frac{1}{R}\left[1 - \frac{1}{2}\left(\frac{2\vec{R}\cdot\vec{a}}{R^2} + \frac{a^2}{R^2}\right) + \frac{3}{8}\left(\frac{2\vec{R}\cdot\vec{a}}{R^2} + \frac{a^2}{R^2}\right)^2 - .....\right]$$

$$= \frac{1}{R}\left[1 - \frac{\vec{a}\cdot\vec{R}}{R^2} - \frac{1}{2}\frac{a^2}{R^2} + \frac{3}{2}\frac{\left(\vec{a}\cdot\vec{R}\right)^2}{R^4} + ...\right]$$

Therefore, we have

$$\frac{1}{\left|\vec{R}+\vec{r}_2\right|} = \frac{1}{R} - \frac{\vec{r}_2\cdot\vec{R}}{R^3} - \frac{1}{2}\frac{r_2^2}{R^3} + \frac{3}{2}\frac{\left(\vec{r}_2\cdot\vec{R}\right)^2}{R^5}$$

$$\frac{1}{\left|\vec{R}-\vec{r}_1\right|} = \frac{1}{R} + \frac{\vec{r}_1\cdot\vec{R}}{R^3} - \frac{1}{2}\frac{r_1^2}{R^3} + \frac{3}{2}\frac{\left(\vec{r}_1\cdot\vec{R}\right)^2}{R^5}$$

$$\frac{1}{\left|\vec{R}+\vec{r}_2-\vec{r}_1\right|} = \frac{1}{R} - \frac{\left(\vec{r}_2-\vec{r}_1\right)\cdot\vec{R}}{R^2} - \frac{1}{2}\frac{\left(\vec{r}_2-\vec{r}_1\right)^2}{R^2} + \frac{3}{2}\frac{\left(\left(\vec{r}_2-\vec{r}_1\right)\cdot\vec{R}\right)^2}{R^4}$$

$$= \frac{1}{R} - \frac{\vec{r}_2\cdot\vec{R}}{R^3} + \frac{\vec{r}_1\cdot\vec{R}}{R^3} - \frac{1}{2}\frac{r_1^2}{R^3} - \frac{1}{2}\frac{r_2^2}{R^3} + \frac{\vec{r}_1\cdot\vec{r}_2}{R^3}$$

$$+ \frac{3}{2}\frac{\left(\vec{r}_2\cdot\vec{R}\right)^2}{R^5} + \frac{3}{2}\frac{\left(\vec{r}_1\cdot\vec{R}\right)^2}{R^5} - 3\frac{\left(\left(\vec{r}_1\cdot\vec{R}\right)\left(\vec{r}_2\cdot\vec{R}\right)\right)}{R^5}$$

Putting all this together we get

$$V = \frac{e^2}{R^3}\left[\vec{r}_1 \cdot \vec{r}_2 - 3\frac{\left((\vec{r}_1 \cdot \vec{R})(\vec{r}_2 \cdot \vec{R})\right)}{R^2}\right]$$

Physically, this says that for large separations, the interaction between the atoms is the same as that between two dipoles $e\vec{r}_1$ and $e\vec{r}_2$ separated by R.

To simplify the algebra, we now choose the vector $\vec{R}$ to lie along the z–axis

$$\vec{R} = R\hat{z}$$

which gives

$$V = \frac{e^2}{R^3}\left[(x_1x_2 + y_1y_2 + z_1z_2) - 3\frac{z_1z_2R^2}{R^2}\right]$$

$$= \frac{e^2}{R^3}(x_1x_2 + y_1y_2 - 2z_1z_2)$$

We now specialize to consider the case $n_1 = n_2 = 2$.

When $n = 2$, there are 4 electron states for each atom

$$\ell = 0, m = 0$$

$$\ell = 1, m = 1, 0, -1$$

Therefore, there are $16 = (4 \times 4)$ degenerate unperturbed zero-order states, each with energy

$$E_0 = -\frac{e^2}{8a_0} - \frac{e^2}{8a_0} = -\frac{e^2}{4a_0}$$

We use degenerate perturbation theory.

To carry out degenerate perturbation theory, we must construct the 16×16 matrix representation of $\langle\, \hat{V}\, \rangle$ and diagonalize it to find the energies corrected to first-order.

The typical matrix element is (leaving off the *n* labels)

$$\begin{aligned}
&\langle \ell_1 m_1 \ell_2 m_2 | \hat{V} | \ell_1 m_1 \ell_2 m_2 \rangle \\
&\quad = \frac{e^2}{R^3} \langle \ell_1 m_1 | \hat{x}_1 | \ell_1 m_1 \rangle \langle \ell_2 m_2 | \hat{x}_2 | \ell_2 m_2 \rangle \\
&\qquad + \frac{e^2}{R^3} \langle \ell_1 m_1 | \hat{y}_1 | \ell_1 m_1 \rangle \langle \ell_2 m_2 | \hat{y}_2 | \ell_2 m_2 \rangle \\
&\qquad - 2\frac{e^2}{R^3} \langle \ell_1 m_1 | \hat{z}_1 | \ell_1 m_1 \rangle \langle \ell_2 m_2 | \hat{z}_2 | \ell_2 m_2 \rangle
\end{aligned}$$

We have

$$x = r\sin\theta\cos\phi = -r\sqrt{\frac{2\pi}{3}}\left(Y_{1,1} - Y_{1,-1}\right)$$

$$y = r\sin\theta\sin\phi = +ir\sqrt{\frac{2\pi}{3}}\left(Y_{1,1} + Y_{1,-1}\right)$$

$$z = r\cos\theta = r\sqrt{\frac{4\pi}{3}}Y_{10}$$

and

$$\begin{aligned}
&\langle n\ell m | x | n\ell' m' \rangle \\
&\quad = -\sqrt{\frac{2\pi}{3}}\left[\int_0^\infty r^3 R_{n\ell}(r) R_{n\ell'}(r) dr\right]\left[\int d\Omega Y^*_{\ell m}(Y_{1,1} - Y_{1,-1}) Y_{\ell' m'}\right] \\
&\langle n\ell m | y | n\ell' m' \rangle \\
&\quad = i\sqrt{\frac{2\pi}{3}}\left[\int_0^\infty r^3 R_{n\ell}(r) R_{n\ell'}(r) dr\right]\left[\int d\Omega Y^*_{\ell m}(Y_{1,1} + Y_{1,-1}) Y_{\ell' m'}\right] \\
&\langle n\ell m | z | n\ell' m' \rangle \\
&\quad = \sqrt{\frac{4\pi}{3}}\left[\int_0^\infty r^3 R_{n\ell}(r) R_{n\ell'}(r) dr\right]\left[\int d\Omega Y^*_{\ell m} Y_{10} Y_{\ell' m'}\right]
\end{aligned}$$

Now let us return to the subject of *selection rules*.

We will just begin the discussion here and then elaborate and finish it later when we cover the topic of time-dependent perturbation theory.

Consider the integrals involving the spherical harmonics above.

We have

$$\int d\Omega Y^*_{\ell m} Y_{1m''} Y_{\ell' m'} = 0 \quad \text{unless} \quad \begin{cases} \ell + \ell' + 1 = even \\ m = m' + m'' \end{cases}$$

These rules follow from doing the integrations over the θ and ϕ variables.

In particular, when the perturbation involves x, y, or z we have

$$\text{for } x \text{ and } y \qquad m = m' \pm 1$$

$$\text{for } z \qquad m = m'$$

which is the so-called

$$\Delta m = \pm 1, 0$$

selection rule for this type of perturbation.

In addition, we have the

$$\Delta \ell = \pm 1$$

selection rule for this type of perturbation.

These two rules will enable us to say many matrix elements are equal to zero by inspection.

We can derive two more very useful selection rules as follows.

We know that

$$\left[\hat{L}_i, r_j\right] = i\hbar\varepsilon_{ijk} r_k$$

This allows us to write (after much algebra)

$$\begin{aligned}\left[\hat{L}_z, \hat{V}\right] &= \left[(\hat{L}_{1z} + \hat{L}_{2z}), \hat{V}\right] = \left[\hat{L}_{1z}, \hat{V}\right] + \left[\hat{L}_{2z}, \hat{V}\right] \\ &= \frac{e^2}{R^3}\left[\hat{L}_{1z}, (x_1x_2 + y_1y_2 - 2z_1z_2)\right] + \frac{e^2}{R^3}\left[\hat{L}_{2z}, (x_1x_2 + y_1y_2 - 2z_1z_2)\right] \\ &= 0\end{aligned}$$

This implies that $[\,\hat{L}_z, \hat{H}\,] = 0$ or that the z–component of the total angular momentum of the electrons is not changed by this perturbation (it is conserved).

This gives the selection rule

$$m_1 + m_2 = m_1' + m_2'$$

Summarizing the selection rules we have

$$\begin{aligned}&\ell_1 + \ell_1' + 1 = even \qquad (= \text{reason b for a zero}) \\ &\ell_2 + \ell_2' + 1 = even \qquad (= \text{reason c for a zero}) \\ &m_1 - m_1' = \pm 1, 0 \qquad (= \text{reason d for a zero}) \\ &m_2 - m_2' = \pm 1, 0 \qquad (= \text{reason d for a zero}) \\ &m_1 + m_2 = m_1' + m_2' \qquad (= \text{reason a for a zero})\end{aligned}$$

and we have also given reason labels for each.

The unperturbed states are (using the format $|\ell_1 m_1\rangle|\ \ell_2 m_2\ \rangle$ are

$$|1\rangle = |00\rangle\,|00\rangle\,,\ |2\rangle = |00\rangle\,|11\rangle\,,\ |3\rangle = |00\rangle\,|10\rangle\,,\ |4\rangle = |00\rangle\,|1,-1\rangle$$
$$|5\rangle = |11\rangle\,|00\rangle\,,\ |6\rangle = |11\rangle\,|11\rangle\,,\ |7\rangle = |11\rangle\,|10\rangle\,,\ |8\rangle = |11\rangle\,|1,-1\rangle$$
$$|9\rangle = |10\rangle\,|00\rangle\,,\ |10\rangle = |10\rangle\,|11\rangle\,,\ |11\rangle = |10\rangle\,|10\rangle\,,\ |12\rangle = |10\rangle\,|1,-1\rangle$$
$$|13\rangle = |1,-1\rangle\,|00\rangle\,,\ |14\rangle = |1,-1\rangle\,|11\rangle\,,\ |15\rangle = |1,-1\rangle\,|10\rangle\,,\ |16\rangle = |1,-1\rangle\,|1,-1\rangle$$

The $\langle\ \hat{V}\ \rangle$ matrix looks like (using labels (VALUE) or (0=reason)) and labeling the rows/columns in order as 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

0bc	0ba	0a	0a	0a	0a	0a	A	0a	0a	B	0a	0a	C	0a	0a
0ba	0bc	0a	0a	D	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a
0a	0a	0a	0bc	0a	0a	0a	0a	E	0a	0a	0a	0a	0a	0a	0a
0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a	0a	F	0a	0a	0a
0a	D	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a
0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a
0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a	0a
A	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a
0a	0a	E	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a
0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a
B	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0bc	0a	0a
0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a
0a	0a	0a	F	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a
C	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0bc	0a	0a
0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a
0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc

$\langle \hat{V} \rangle$ matrix entries

There are only 12 nonzero elements(out of 256) and because the matrix is Hermitian we only have 6 to calculate (one side of the diagonal).

It should now be clear why finding the relevant selection rules is so important!!!

The nonzero elements are given by the expressions

$$A = C = \frac{e^2}{R^3} \langle 200| \langle 200| (x_1 x_2 + y_1 y_2) |211\rangle |21, -1\rangle$$

$$B = E = -2\frac{e^2}{R^3} \langle 200| \langle 200| z_1 z_2 |210\rangle |21, 0\rangle$$

$$D = \frac{e^2}{R^3} \langle 200| \langle 211| (x_1 x_2 + y_1 y_2) |211\rangle |200\rangle$$

$$F = \frac{e^2}{R^3} \langle 200| \langle 21, -1| (x_1 x_2 + y_1 y_2) |21, -1\rangle |200\rangle$$

If we define

$$\alpha = \sqrt{\frac{8\pi}{3}} \int_0^\infty r^3 R_{20} R_{21} dr$$

we have

$$\langle 200| x |211\rangle = \frac{\alpha}{2} = -\langle 200| x |21, -1\rangle$$

$$\langle 200| y |211\rangle = i\frac{\alpha}{2} = \langle 200| x |21, -1\rangle$$

$$\langle 200| z |211\rangle = \frac{\sqrt{2}}{2}\alpha$$

and

$$A = C = \frac{E}{2} = \frac{B}{2} = -D = -F = -\frac{1}{2}\alpha^2 \frac{e^2}{R^3}$$

Now we rearrange the row/column labels(the original choice was arbitrary) to create a *Jordan canonical form* with blocks on the diagonal.

We choose

1 8 11 14 2 5 3 9 4 13 6 7 10 12 15 16

0	A	2A	A	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	-A	0	0	0	0	0	0	0	0	0	0
0	0	0	0	-A	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2A	0	0	0	0	0	0	0	0
0	0	0	0	0	0	2A	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	-A	0	0	0	0	0
0	0	0	0	0	0	0	0	0	-A	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Jordan Canonical Form

The is called the *block-diagonalized form*.

We have one 4×4 and three 2×2 matrices to diagonalize.

We get the eigenvalues

$$4 \times 4 \rightarrow 0, 0, \pm\sqrt{6}A \quad , \quad 2 \times 2 \rightarrow \pm A$$
$$2 \times 2 \rightarrow \pm 2A \quad , \quad 2 \times 2 \rightarrow \pm A$$

Therefore, the energies correct to first-order are

$$E = \begin{cases} E_0 + \sqrt{6}A & \text{degeneracy} = 1 \\ E_0 + 2A & \text{degeneracy} = 1 \\ E_0 + A & \text{degeneracy} = 2 \\ E_0 & \text{degeneracy} = 8 \\ E_0 - A & \text{degeneracy} = 2 \\ E_0 - 2A & \text{degeneracy} = 1 \\ E_0 - \sqrt{6}A & \text{degeneracy} = 1 \end{cases}$$

That was a **real** problem!

Variational Methods

All perturbation methods rely on our ability to make the separation $\hat{H} = \hat{H}_0 + \hat{V}$ where $\hat{H}_0$ is solvable exactly and $\hat{V}$ is a small correction.

The Rayleigh-Ritz variational method is not subject to any such restrictions.

This method is based on the following mathematical results.

We can always write

$$\hat{H} = \hat{I}\hat{H} = \sum_N |N\rangle \langle N| \hat{H} = \sum_N E_n |N\rangle \langle N|$$

where

$$\hat{H} |N\rangle = E_n |N\rangle$$

This is just the spectral decomposition of $\hat{H}$ in terms of its eigenvectors and eigenvalues.

Now, if we choose some arbitrary state vector $|\psi\rangle$ (called a trial vector), then have

$$\begin{aligned}
\langle\psi| \hat{H} |\psi\rangle &= \sum_N E_n \langle\psi \mid N\rangle \langle N \mid \psi\rangle \\
&\geq \sum_N E_0 \langle\psi \mid N\rangle \langle N \mid \psi\rangle = E_0 \sum_N \langle\psi \mid N\rangle \langle N \mid \psi\rangle \\
&\geq E_0 \langle\psi| \left(\sum_N |N\rangle \langle N| \right) |\psi\rangle = E_0 \langle\psi| \hat{I} |\psi\rangle = E_0 \langle\psi \mid \psi\rangle
\end{aligned}$$

or

$$\frac{\langle\psi| \hat{H} |\psi\rangle}{\langle\psi \mid \psi\rangle} \geq E_0$$

for *any choice* of the trial vector $|\psi\rangle$, where E_0 is the ground state energy (the lowest energy).

Equality holds only if $|\psi\rangle$ is the true ground state vector.

This result says that

$$\frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi\,|\,\psi\rangle} \text{ is an upper bound for } E_0$$

Procedure

1. Pick a trial vector $|\psi\rangle$ that contains unknown parameters $\{\alpha_k\}$
2. Calculate
$$\frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi\,|\,\psi\rangle} = E_0\left(\{\alpha_k\}\right)$$
3. Since $E_0\left(\{\alpha_k\}\right)$ is an upper bound, we then minimize it with respect to all of the parameters $\{\alpha_k\}$. This gives a least upper bound for that choice of the functional form for the trial vector.
4. We perform the minimization by setting
$$\frac{\partial E_0}{\partial \alpha_k} = 0 \text{ for all } k$$
5. The more complex the trial vector, i.e., the more parameters we incorporate allows us to better approximate the true functional form of the ground state vector and we will get closer and closer to the true energy value.

What about states other than the ground state?

If the ground state has different symmetry properties than the first excited state, i.e.,

$$\text{ground state} \rightarrow \ell = 0 \rightarrow \text{ contains } Y_{00}$$

$$1^{st} \text{ excited state} \rightarrow \ell = 1 \rightarrow \text{ contains } Y_{1m}$$

then if we choose a trial vector with the symmetry of an $\ell = 0$ state we obtain an approximation to the ground state energy.

If, however, we choose a trial vector with the symmetry of an $\ell = 1$ state, then we obtain an approximation to the first-excited state energy and so on.

In other words, the variational method always gives the least upper bound for the energy of the state with the *same symmetry* as the trial vector.

Example

Let us choose the harmonic oscillator Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}kx^2$$

and a trial wave function

$$\psi(x,a) = \begin{cases} (a^2 - x^2)^2 & |x| < a \\ 0 & |x| \geq a \end{cases}$$

where *a* is an unknown parameter.

The variational principle says that

$$\frac{\langle\psi(a)|\,\hat{H}\,|\psi(a)\rangle}{\langle\psi(a)\,|\,\psi(a)\rangle} = E_0(a) \geq E_0 = \text{ true ground state energy}$$

We get a *best* value for this choice of trial function by minimizing with respect to a using

$$\frac{dE_0(a)}{da} = 0$$

Now we need to calculate the integrals.

We have for the denominator (the normalization integral)

$$\langle\psi(a)\,|\,\psi(a)\rangle = \int_{-a}^{a} \psi^2(x,a)dx = \int_{-a}^{a} (a^2 - x^2)^4 dx$$

$$= 2\int_{0}^{a} (a^2 - x^2)^4 dx = \frac{336}{315}a^9$$

and for the numerator

$$\langle\psi(a)|\,\hat{H}\,|\psi(a)\rangle = -\frac{\hbar^2}{2m}2\int_{0}^{a} \psi(x,a)\frac{d^2\psi(x,a)}{dx^2}dx + \frac{1}{2}k2\int_{0}^{a} \psi^2(x,a)x^2dx$$

$$= -\frac{\hbar^2}{2m}\left(-\frac{80}{21}a^7\right) + \frac{1}{2}k\left(\frac{336}{3465}a^{11}\right)$$

Therefore

$$E_0(a) = 1.786\frac{\hbar^2}{ma^2} + 0.045ka^2$$

The minimum condition gives

$$a^2 = 6.300\left(\frac{\hbar^2}{mk}\right)^{1/2}$$

which then says that

$$0.566\hbar\omega \geq E_0$$

The true value is $0.500\hbar\omega$.

This is an excellent result considering that the trial function does not look at all like the correct ground-state wave function (it is a Gaussian function).

This points out clearly how powerful the variational technique can be for many problems.

Chapter 9

Time-Dependent Perturbation Theory

Theory

Time-independent or stationary-state perturbation theory, as we developed earlier, allows us to find approximations for the energy eigenvalues and eigenvectors in complex physical systems that are not solvable in closed form and where we could write $\hat{H}$ in two parts as

$$\hat{H} = \hat{H}_0 + \hat{V}$$

For these perturbation methods to work, $\hat{V}$ must be *weak* and *time-independent*.

We now turn our attention to the case

$$H = H_0 + V_t$$

where $\hat{V}_t$ is *weak* and *time-dependent*.

Examples might be the decays of an atomic system by photon emission or the ionization of an atom by shining light on it.

We assume that at some time t_0 the system has evolved into the state $\left|\psi_t^{(0)}\right\rangle$, i.e., the state $\left|\psi_t^{(0)}\right\rangle$ satisfies the time evolution equation

$$i\hbar\frac{\partial}{\partial t}\left|\psi_t^{(0)}\right\rangle = \hat{H}_0\left|\psi_t^{(0)}\right\rangle \qquad t \le t_0$$

It is a solution of the time-dependent Schrodinger equation with no perturbing interactions before t_0 where

$$\hat{H} = \hat{H}_0 \qquad t \le t_0$$

At time t_0 we *turn on* the interaction potential (or perturbation) so that

$$\hat{H} = \hat{H}_0 + \hat{V}_t \qquad t \ge t_0$$

The new state of the system then satisfies

$$i\hbar \frac{\partial}{\partial t} |\psi_t\rangle = \hat{H} |\psi_t\rangle = (\hat{H}_0 + \hat{V}_t) |\psi_t\rangle \qquad t \ge t_0$$

with the *boundary condition*(*initial value*)

$$|\psi_t\rangle = \left|\psi_t^{(0)}\right\rangle \text{ at } t{=}t_0$$

As we said, we assume that the full time-dependent Schrodinger equation cannot be solved in closed form and so we look for approximate solutions.

We let $\hat{V}_t$ be a small perturbation, i.e., we assume there is a natural small parameter in $\hat{V}_t$ (as we saw in time-independent perturbation theory) and we make an expansion of the solution in powers of $\hat{V}_t$ or this small parameter.

Since the effect of $\hat{H}_0$ will be much greater than the effect of $\hat{V}_t$, most of the time dependence comes from $\hat{H}_0$.

If we could neglect $\hat{V}_t$, then since $\hat{H}_0$ is independent of time, we would have the simple time dependence

$$|\psi_t\rangle = e^{-\frac{i}{\hbar}\hat{H}_0 t} \left|\psi_t^{(0)}\right\rangle$$

Let us assume that this is still approximately true and remove this known time dependence from the solution.

This should remove the major portion of the total time dependence from the problem.

We do this by assuming a solution of the form

$$|\psi_t\rangle = e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|\psi(t)\rangle$$

and then determining and solving the equation for the new state vector $|\psi(t)\rangle$.

Substituting this assumption in our original equation, the equation for $|\psi(t)\rangle$ is then given by

$$i\hbar\frac{\partial}{\partial t}\left(e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|\psi(t)\rangle\right) = (\hat{H}_0 + \hat{V}_t)\left(e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|\psi(t)\rangle\right)$$

$$\hat{H}_0 e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|\psi(t)\rangle + i\hbar e^{-\frac{i}{\hbar}\hat{H}_0 t}\frac{\partial}{\partial t}\,|\psi(t)\rangle$$

$$= \hat{H}_0 e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|\psi(t)\rangle + \hat{V}_t e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|\psi(t)\rangle$$

$$i\hbar\frac{\partial}{\partial t}\,|\psi(t)\rangle = \hat{V}(t)\,|\psi(t)\rangle$$

where

$$\hat{V}(t) = e^{\frac{i}{\hbar}\hat{H}_0 t}\hat{V}_t e^{-\frac{i}{\hbar}\hat{H}_0 t}$$

The substitution has removed $\hat{H}_0$ from the equation and changed the time dependence of the perturbing potential.

We are in the so-called *interaction picture* or *representation* where *both* the state vectors and the operators depend on time.

We develop a *formal solution* by integrating this equation of motion for the state vector to get

$$i\hbar \int_{t_0}^{t} \frac{\partial}{\partial t'} |\psi(t')\rangle dt' = \int_{t_0}^{t} \hat{V}(t') |\psi(t')\rangle dt'$$
$$= i\hbar(|\psi(t)\rangle - |\psi(t_0)\rangle)$$

so that the formal solution is given by

$$|\psi(t)\rangle = |\psi(t_0)\rangle + \frac{1}{i\hbar} \int_{t_0}^{t} \hat{V}(t') |\psi(t')\rangle dt'$$

This is an *integral equation* for $|\psi(t)\rangle$.

We solve it as a power series in $\hat{V}_t$ by the *method of iteration*.

The 0^{th}–order approximation is found by neglecting the perturbing potential.

We get $$|\psi(t)\rangle = |\psi(t_0)\rangle$$

The 1^{st}–order approximation is obtained by inserting the 0^{th}–order approximation into the full equation.

We get

$$|\psi(t)\rangle = |\psi(t_0)\rangle + \frac{1}{i\hbar} \int_{t_0}^{t} \hat{V}(t') |\psi(t_0)\rangle dt'$$
$$= \left(1 + \frac{1}{i\hbar} \int_{t_0}^{t} \hat{V}(t') dt'\right) |\psi(t_0)\rangle$$

The 2nd–order approximation is obtained by inserting the 1st–order approximation into the full equation.

We get

$$|\psi(t)\rangle = |\psi(t_0)\rangle + \frac{1}{i\hbar}\int_{t_0}^{t} \hat{V}(t')\,|\psi(t_0)\rangle dt'$$

$$+ \frac{1}{(i\hbar)^2}\int_{t_0}^{t} dt' \int_{t_0}^{t'} dt'' \hat{V}(t')\hat{V}(t'')\,|\psi(t_0)\rangle$$

Notice that in all subsequent iterations the operators $\hat{V}(t'), \hat{V}(t'')$,. . ., etc, always occur in order of increasing time from right to left.

We can write the general result as

$$|\psi(t)\rangle = \hat{U}(t, t_0)\,|\psi(t_0)\rangle$$

where

$$\hat{U}(t, t_0) = \hat{I} + \sum_{n=1}^{\infty} \frac{1}{(i\hbar)^n} \int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 \ldots\ldots\ldots \int_{t_0}^{t_{n-1}} dt_n \hat{V}(t_1)\hat{V}(t_2)\ldots\ldots\hat{V}(t_n)$$

The complete, formal solution to the problem is then given by

$$|\psi_t\rangle = e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|\psi(t)\rangle = e^{-\frac{i}{\hbar}\hat{H}_0 t}\hat{U}(t, t_0)\,|\psi(t_0)\rangle$$

so that

$$e^{-\frac{i}{\hbar}\hat{H}_0 t}\hat{U}(t, t_0) = \text{ the total time development operator}$$

Before developing the detailed techniques of time-dependent perturbation theory, let us spend some time with the operator $\hat{U}(t, t_0)$ and discuss some of its properties.

We first introduce the idea of a *time-ordered product* of operators as follows.

The symbol

$$\left(\hat{A}(t)\hat{B}(t_1)\hat{C}(t_2).....\hat{X}(t_n)\right)_+ \equiv T\left(\hat{A}(t)\hat{B}(t_1)\hat{C}(t_2).....\hat{X}(t_n)\right)$$

means the product of the operators where the operators are written from right to left in order of increasing times, i.e.,

$$\left(\hat{A}(t)\hat{B}(t')\right)_+ = \begin{cases} \hat{A}(t)\hat{B}(t') & t \geq t' \\ \hat{B}(t')\hat{A}(t) & t' \geq t \end{cases}$$

Now, we have using the time-ordered product definition

$$\left(\left[\int_{t_0}^{t} \hat{V}(t')dt'\right]^2\right)_+ = \left(\int_{t_0}^{t} \hat{V}(t')dt' \int_{t_0}^{t} \hat{V}(t'')dt''\right)_+$$

$$= \int_{t_0}^{t} dt' \int_{t_0}^{t} dt''(\hat{V}(t')\hat{V}(t''))_+$$

$$= \int_{t_0}^{t} dt' \int_{t_0}^{t'} dt''\hat{V}(t')\hat{V}(t'') + \int_{t_0}^{t} dt'' \int_{t_0}^{t''} dt'\hat{V}(t'')\hat{V}(t')$$

$$= 2\int_{t_0}^{t} dt' \int_{t_0}^{t'} dt''\hat{V}(t')\hat{V}(t'')$$

and in general

$$\left(\left[\int_{t_0}^{t}\hat{V}(t')dt'\right]^n\right)_+ = \left(\int_{t_0}^{t}\hat{V}(t_1)dt_1\int_{t_0}^{t}\hat{V}(t_2)dt_2.....\int_{t_0}^{t}\hat{V}(t_n)dt_n\right)_+$$

$$= \int_{t_0}^{t}dt_1\int_{t_0}^{t}dt_2.....\int_{t_0}^{t}dt_n(\hat{V}(t_1)\hat{V}(t_2).....\hat{V}(t_n))_+$$

$$= n!\int_{t_0}^{t}dt_1\int_{t_0}^{t}dt_2.....\int_{t_0}^{t}dt_n\hat{V}(t_1)\hat{V}(t_2).....\hat{V}(t_n)$$

because there are $n!$ possible orderings of the n terms involved.

This last form is identical to the expression for $\hat{U}(t,t_0)$ and thus we have

$$\hat{U}(t,t_0) = \sum_{n=0}^{\infty}\frac{1}{(i\hbar)^n}\frac{1}{n!}\left(\left[\int_{t_0}^{t}\hat{V}(t')dt'\right]^n\right)_+$$

$$= \left(\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t}\hat{V}(t')dt'\right]\right)_+$$

The last expression is just a *convenient shorthand for the infinite sum*.

In order to verify that this is in fact a solution of

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = \hat{V}(t)|\psi(t)\rangle$$

we must prove that

$$i\hbar\frac{\partial}{\partial t}\hat{U}(t,t_0)\,|\psi(t_0)\rangle = \hat{V}(t)\hat{U}(t,t_0)\,|\psi(t_0)\rangle$$

$$i\hbar\frac{\partial}{\partial t}\hat{U}(t,t_0) = \hat{V}(t)\hat{U}(t,t_0)$$

Substituting, we have

$$i\hbar\frac{\partial}{\partial t}\hat{U}(t,t_0) = i\hbar\frac{\partial}{\partial t}\left(\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t}\hat{V}(t')dt'\right]\right)_+$$

$$= \left(\hat{V}(t)\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t}\hat{V}(t')dt'\right]\right)_+$$

In the differentiation we do not have to worry about the non-commutation of the operators inside the time-ordered product since the order is already specified.

Since t is certainly the latest time in the time-ordered product and therefore all the other operators will be on the right of $\hat{V}(t)$ we can pull it outside the time-ordered product and write

$$i\hbar\frac{\partial}{\partial t}\hat{U}(t,t_0) = \hat{V}(t)\left(\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t}\hat{V}(t')dt'\right]\right)_+ = \hat{V}(t)\hat{U}(t,t_0)$$

as required.

The *most important question* (really the only question) that is usually asked in quantum mechanics is the following:

> Suppose that the system is initially in an eigenstate$|n\rangle$ of $\hat{H}_0$, i.e., $\hat{H}_0\,|n\rangle = \varepsilon_n|n\rangle$.
> What is the probability that the system will be observed, after the perturbation has had time to act, in a different (and thus orthogonal) eigenstate of $\hat{H}_0$, say $|m\rangle$?

Alternatively, the question is sometimes posed this way:

> What is the probability that the interaction causes the system to make a *transition* from the state $|n\rangle$ to the state $|m\rangle$?

The probability amplitude for observing the system in the state $|m\rangle$ at time t is given by

$$\begin{aligned}\langle m \mid \psi_t \rangle &= \langle m| e^{-\frac{i}{\hbar}\hat{H}_0 t}\hat{U}(t,t_0)\,|\psi(t_0)\rangle \\ &= \langle m| e^{-\frac{i}{\hbar}\hat{H}_0 t}\hat{U}(t,t_0)\,|n\rangle\end{aligned}$$

where

$$|\psi(t_0)\rangle = |n\rangle$$

is the initial state.

Setting $t_0 = 0$ for simplicity and using the 1st–order approximation for $\hat{U}(t,0)$ and also using

$$\langle m| e^{-\frac{i}{\hbar}\hat{H}_0 t} = \left(e^{\frac{i}{\hbar}\hat{H}_0 t}\,|m\rangle\right)^{\dagger} = \left(e^{\frac{i}{\hbar}\varepsilon_m t}\,|m\rangle\right)^{\dagger} = \langle m| e^{-\frac{i}{\hbar}\varepsilon_m t}$$

we get

$$\begin{aligned}\langle m \mid \psi_t \rangle &= \frac{1}{i\hbar} e^{-\frac{i}{\hbar}\varepsilon_m t}\int_0^t dt'\,\langle m|\hat{V}(t')\,|n\rangle \\ &= \frac{1}{i\hbar} e^{-\frac{i}{\hbar}\varepsilon_m t}\int_0^t dt'\,\langle m| e^{\frac{i}{\hbar}\hat{H}_0 t}\hat{V}_{t'} e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|n\rangle \\ &= \frac{1}{i\hbar} e^{-\frac{i}{\hbar}\varepsilon_m t}\int_0^t dt'\, e^{\frac{i}{\hbar}(\varepsilon_m-\varepsilon_n)t}\,\langle m|\hat{V}_{t'}\,|n\rangle\end{aligned}$$

The probability of the transition is then

$$P_{n\to m}(t) = |\langle m \mid \psi_t\rangle|^2 = \frac{1}{\hbar^2}\left|\int_0^t dt' e^{\frac{i}{\hbar}(\varepsilon_m-\varepsilon_n)t}\,\langle m|\,\hat{V}_{t'}\,|n\rangle\right|^2$$

The simplest example is when $\hat{V}_t$ is not a function of t, or $\hat{V}_t = \hat{V}$.

We then have

$$P_{n\to m}(t) = |\langle m \mid \psi_t\rangle|^2 = \frac{\left|\langle m|\,\hat{V}_{t'}\,|n\rangle\right|^2}{\hbar^2}\left|\int_0^t dt' e^{\frac{i}{\hbar}(\varepsilon_m-\varepsilon_n)t}\right|^2$$

If we define $\Delta = \varepsilon_m - \varepsilon_n$, then we have

$$P_{n\to m}(t) = \left|\langle m|\,\hat{V}_{t'}\,|n\rangle\right|^2\left|\frac{1-e^{-\frac{i}{\hbar}\Delta t}}{\Delta}\right|^2 = \left|\langle m|\,\hat{V}_{t'}\,|n\rangle\right|^2\left(\frac{\sin\frac{\Delta t}{2\hbar}}{\Delta/2}\right)^2$$

for the transition probability.

What is the physical meaning of this result?

We must be *very careful* when we use the words

> the perturbation causes a transition
> between eigenstates of $\hat{H}_0$

What this means physically is that the system has absorbed from the perturbing field (or emitted to it) the energy difference $\Delta = \varepsilon_m - \varepsilon_n$ and therefore the system has changed its energy.

Does the statement also mean that the state vector has changed from an initial value $|\psi(0)\rangle = |n\rangle$ to a final value $|\psi(t)\rangle = |m\rangle$?

We can get a better feeling for the correct answer to this question by deriving the result in a different manner.

We have

$$i\hbar\frac{\partial}{\partial t}|\psi_t\rangle = \hat{H}|\psi_t\rangle = (\hat{H}_0 + \hat{V}_t)|\psi_t\rangle$$

and

$$\hat{H}_0|n\rangle = \varepsilon_n|n\rangle$$

As in our development of time-independent perturbation theory, we let

$$\hat{V}_t = g\hat{U}_t$$

where g is a small parameter.

The set of eigenvectors $\{|n\rangle\}$ is a complete set and therefore we can use it as a basis for the space and, in particular, we can write

$$|\psi_t\rangle = \sum_n a_n(t)e^{-\frac{i}{\hbar}\varepsilon_n t}|n\rangle$$

The reason for pulling out the phase factors will be clear shortly.

It is clear that if $g = 0$, then this is the correct general solution with

$$a_n(t) = a_n(0) = \text{constant}$$

The phase factors we pulled out represent the time dependence due to $\hat{H}_0$ and this is, by assumption, the major time dependence in the system.

If g is small we expect the time dependence of $a_n(t)$, which is due to the perturbation to be weak or that

$$\frac{da_n(t)}{dt} \text{ is small}$$

It is in this sense that we can propose to use perturbation theory on the system.

Using the eigenbasis expansion we have

$$\sum_n \left(i\hbar \frac{da_n(t)}{dt} + \varepsilon_n a_n(t) \right) e^{-\frac{i}{\hbar}\varepsilon_n t} |n\rangle$$
$$= \sum_n \left(\varepsilon_n a_n(t) + g\hat{U}_t a_n(t) \right) e^{-\frac{i}{\hbar}\varepsilon_n t} |n\rangle$$

Applying the linear functional $\langle m|$ from the left and using the orthonormality relation

$$\langle m \mid n \rangle = \delta_{mn}$$

we get

$$i\hbar \frac{da_m(t)}{dt} = g \sum_n \langle m| \hat{U}_t |n\rangle e^{i\omega_{mn}t} a_n(t)$$
$$= \sum_n V_{mn}(t) e^{i\omega_{mn}t} a_n(t)$$

where

$$\omega_{mn} = \frac{\varepsilon_m - \varepsilon_n}{\hbar}$$

This is an exact equation.

It implies that the time dependence of $a_n(t)$ is due entirely to $\hat{V}_t$ (because we explicitly extracted out the dependence due to $\hat{H}_0$).

This is the interaction picture that we had earlier.

Exactly Solvable 2-State Example

Consider a 2-state system with

$$H_0 = \begin{pmatrix} E_1 & 0 \\ 0 & E_2 \end{pmatrix} \quad , \quad V(t) = \begin{pmatrix} 0 & \delta e^{i\omega t} \\ \delta e^{-i\omega t} & 0 \end{pmatrix} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

In the interaction picture, as derived above, we have

$$i\hbar \frac{dc_m(t)}{dt} = \sum_n V_{mn}(t) e^{i\omega_{mn} t} c_n(t) \quad , \quad |\psi_t\rangle = \sum_n c_n(t) e^{-\frac{i}{\hbar}\varepsilon_n t} |n\rangle$$

or

$$i\hbar \frac{dc_1(t)}{dt} = \delta e^{i\left[\omega + \frac{E_1 - E_2}{\hbar}\right]t} c_2(t)$$

$$i\hbar \frac{dc_2(t)}{dt} = \delta e^{i\left[-\omega - \frac{E_1 - E_2}{\hbar}\right]t} c_1(t)$$

$$|\psi_t\rangle = c_1(t) e^{-\frac{i}{\hbar}E_1 t} |1\rangle + c_2(t) e^{-\frac{i}{\hbar}E_2 t} |2\rangle$$

We can write these equations as

$$\frac{dC(t)}{dt} = -\frac{i\delta}{\hbar} \begin{pmatrix} 0 & e^{i[\omega - \omega_{21}]t} \\ e^{-i[\omega - \omega_{21}]t} & 0 \end{pmatrix} C(t)$$

where

$$C(t) = \begin{pmatrix} c_1(t) \\ c_2(t) \end{pmatrix} \quad , \quad \omega_{21} = \frac{E_2 - E_1}{\hbar}$$

We can find an exact solution.

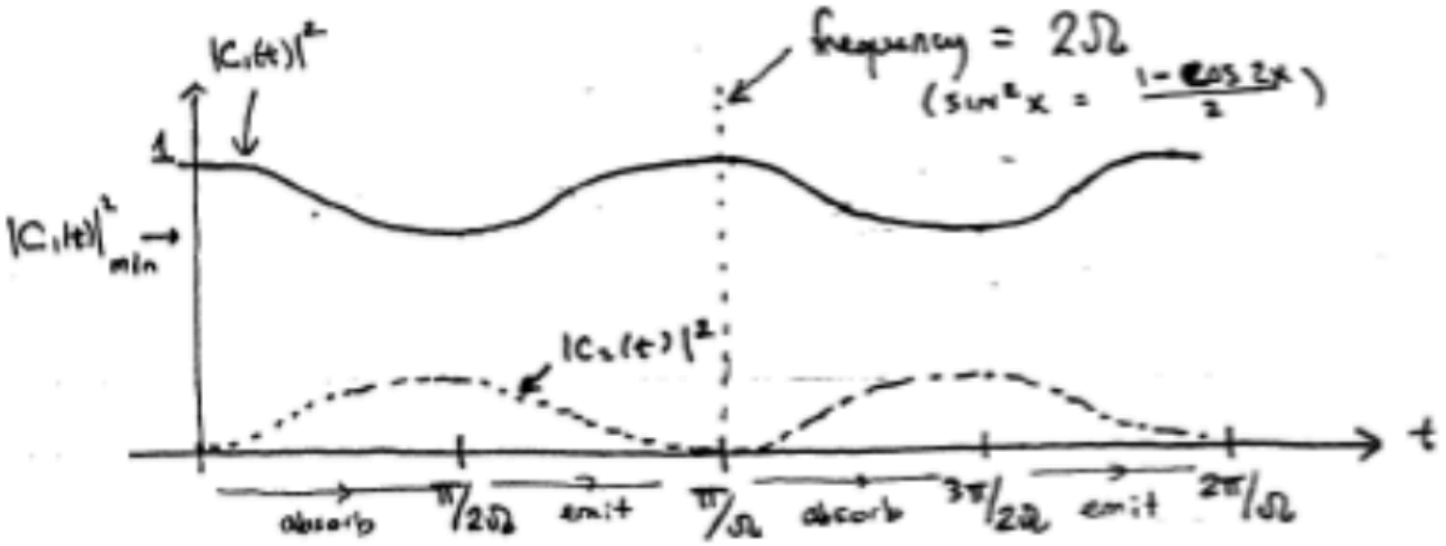
With initial conditions $c_1(0) = 1$ and $c_2(0) = 0$ we get

$$|c_1(t)|^2 = \frac{\delta^2}{\delta^2 + \frac{\hbar^2(\omega-\omega_{21})^2}{4}} \sin^2 \Omega t = 1 - |c_2(t)|^2$$

where

$$\Omega^2 = \frac{\delta^2}{\hbar^2} + \frac{(\omega - \omega_{21})^2}{4}$$

A graph of these functions is shown in the Figure below.



Exact Solution

A straightforward calculation gives

$$|c_1(t)|^2_{\min} = \frac{(\omega - \omega_{21})^2}{(\omega - \omega_{21})^2 + \frac{4\delta^2}{\hbar^2}}$$

At resonance, $\omega = \omega_{21}$, we have

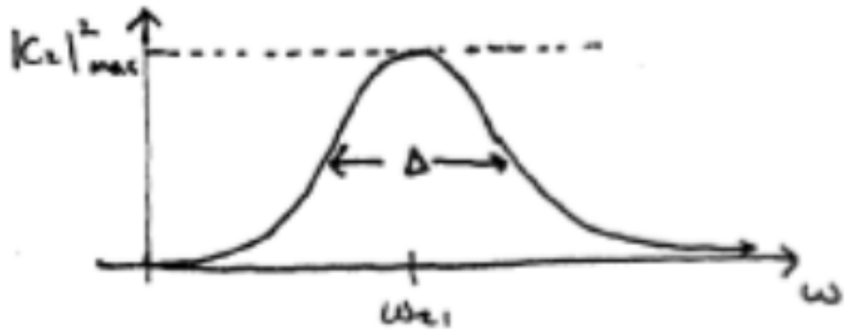
$$\Omega = \frac{\delta}{\hbar} \quad , \quad |c_1(t)|^2_{\min} = 0$$

as shown in the Figure.



At Resonance

The amplitude as a function of ω is shown in the Figure below.



Amplitude versus ω

where Δ = full width at half maximum = $4\delta/\hbar$.

The amplitude is peaked at resonance and the width is proportional to δ (the strength of the perturbation).

This periodically forced 2–state system is a basic problem - it demonstrates the fundamental features of absorption and emission.

We now return to the full, general equations and look for a perturbation solution.

Now we assume (power series)

$$a_n(t) = a_n^{(0)} + g a_n^{(1)} + g^2 a_n^{(2)} + \ldots\ldots$$

Substituting and arranging the terms in a power series in g we have

$$\left(\frac{da_n^{(0)}}{dt}\right) g^0 + \left(i\hbar \frac{da_m^{(1)}}{dt} - \sum_n \langle m|\hat{U}_t|n\rangle e^{\frac{i}{\hbar}\omega_{mn}t} a_n^{(0)}\right) g^1 + \ldots\ldots +$$

$$\left(i\hbar \frac{da_m^{(r+1)}}{dt} - \sum_n \langle m|\hat{U}_t|n\rangle e^{\frac{i}{\hbar}\omega_{mn}t} a_n^{(r)}\right) g^r + \ldots\ldots = 0$$

or looking at each order separately we have

$$0^{th}-\text{order} \quad \frac{da_n^{(0)}}{dt}=0$$

$$1^{st}-\text{order} \quad i\hbar\frac{da_m^{(1)}}{dt}=\sum_n \langle m|\hat{U}_t|n\rangle e^{\frac{i}{\hbar}\omega_{mn}t}a_n^{(0)}$$

.....

.....

$$(r+1)^{st}-\text{order} \quad i\hbar\frac{da_m^{(r+1)}}{dt}=\sum_n \langle m|\hat{U}_t|n\rangle e^{\frac{i}{\hbar}\omega_{mn}t}a_n^{(r)}$$

.....

Note that the coefficients $a_n^{(0)}$ follow from the initial condition

$$|\psi(0)\rangle=\sum_n a_n^{(0)}|n\rangle$$

The solution proceeds as follows:

$$\text{initial condition} \rightarrow a_n^{(0)}$$

$$a_n^{(0)} \rightarrow a_n^{(1)} \text{ using the } 1^{st}-\text{order equation}$$

.....

$$a_n^{(r)} \rightarrow a_n^{(r+1)} \text{ using the } (r+1)^{st}-\text{order equation}$$

Now consider the following example.

We assume that

$$\hat{H}=\hat{H}_0 \qquad t\le 0$$

where

$$\hat{H}_0|n\rangle=\varepsilon_n|n\rangle$$

and during the time interval $0 \le t \le T$ a perturbation $\hat{V}_t$ is applied to the system and the $a_n(t)$ change with time.

Finally, for $t \ge T$ the perturbation is turned off and $a_n(t) = a_n(T)$.

The probability that, as a result of the perturbation, the energy of the system becomes ε_r, is given by

$$|\langle r \mid \psi_t \rangle|^2 = \left| \sum_i a_i(t) e^{-\frac{i}{\hbar}\varepsilon_i t} \langle r \mid i \rangle \right|^2 = |a_r(t)|^2$$

and as $t \to \infty$ we get

$$|\langle r \mid \psi_t \rangle|^2 = |a_r(T)|^2$$

Now to 1st–order we have

$$i\hbar \frac{da_r^{(1)}}{dt} = \sum_n \langle r| \hat{U}_t |n\rangle e^{\frac{i}{\hbar}\omega_{rn} t} a_n^{(0)}$$

If $|\psi(0)\rangle = |i\rangle$, then

$$a_n^{(0)} = \begin{cases} 1 & n = i \\ 0 & n \neq i \end{cases}$$

This gives

$$i\hbar \frac{da_r^{(1)}}{dt} = \langle r| \hat{U}_t |i\rangle e^{\frac{i}{\hbar}\omega_{ri} t}$$

Integrating we have

$$a_r^{(1)}(T) = \frac{1}{i\hbar} \int_0^T \langle r| \hat{U}_t |i\rangle e^{\frac{i}{\hbar}(\varepsilon_r - \varepsilon_i)t} dt$$

and

$$a_r(T) = a_r^{(0)}(T) + g a_r^{(1)}(T)$$

which is identical to our earlier result as $t \to \infty$.

Now let return to our question.

Has the state changed also?

In the example we found that the perturbation produces a final state $|\psi_t\rangle$ for $t \geq T$ which to 1st–order is

$$|\psi_t\rangle = \sum_n a_n(t) e^{-\frac{i}{\hbar}\varepsilon_n t} |n\rangle$$

This is a coherent (definite relative phases) superposition of eigenvectors of $\hat{H}_0$.

This is NOT a stationary state.

Interference effects between the terms in the sum are detectable.

They do not, however, affect

$$|a_r(T)|^2 = \text{ probability that the energy changes to } \varepsilon_r$$

Thus, the perturbation does not cause a jump from one stationary state $|i\rangle$ of $\hat{H}_0$ to another $|r\rangle$, but instead it produces a non-stationary state.

The conventional language of quantum mechanics produces this ambiguity between the two statements

the energy is ϵ_r and *the state is* $|r\rangle$

For the state

$$|\psi_t\rangle = \sum_n a_n(t) e^{-\frac{i}{\hbar}\varepsilon_n t} |n\rangle$$

it is correct to say

the probability of the energy being ϵ_r is $|a_r(T)|^2$

or

$$Prob(E = \epsilon_r | |\psi\rangle) = |a_r(T)|^2$$

The state, however, is $|\psi_t\rangle$ and NOT $|r\rangle$.

An example

Suppose we perturb an oscillator with a decaying electric field of the form

$$\hat{V}_t = -q\mathcal{E}_0\hat{x}e^{-\frac{t}{\tau}} \qquad t \geq 0$$

To 1st–order, starting with the initial state $|n\rangle$ with energy

$$\varepsilon_n = \hbar\omega(n + \frac{1}{2})$$

we have

$$|\psi(t)\rangle = |n\rangle + \frac{1}{i\hbar}\int_0^t dt'\hat{V}(t)\,|n\rangle$$

where

$$\hat{V}(t) = e^{\frac{i}{\hbar}\hat{H}_0 t}\hat{V}_t e^{-\frac{i}{\hbar}\hat{H}_0 t}$$

We let $n = 0$ (the ground state) for this example.

We then have

$$|\psi(t)\rangle = |0\rangle + \frac{1}{i\hbar}\int_0^t dt' e^{\frac{i}{\hbar}\hat{H}_0 t}(-q\mathcal{E}_0\hat{x})e^{-\frac{t}{\tau}}e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|0\rangle$$

Using

$$e^{-\frac{i}{\hbar}\hat{H}_0 t}\,|0\rangle = e^{-i\frac{\omega}{2}t}\,|0\rangle$$

$$\hat{x}\,|0\rangle = \sqrt{\frac{\hbar}{2m\omega}}\,|1\rangle$$

$$e^{\frac{i}{\hbar}\hat{H}_0 t}\,|1\rangle = e^{i\frac{3\omega}{2}t}\,|0\rangle$$

we get (letting $t \to \infty$)

$$|\psi(t)\rangle = |0\rangle + \frac{1}{i\hbar}(-q\mathcal{E}_0)\sqrt{\frac{\hbar}{2m\omega}}\int_0^\infty dt' e^{i\omega t' - \frac{t}{\tau}}|1\rangle$$

and finally,

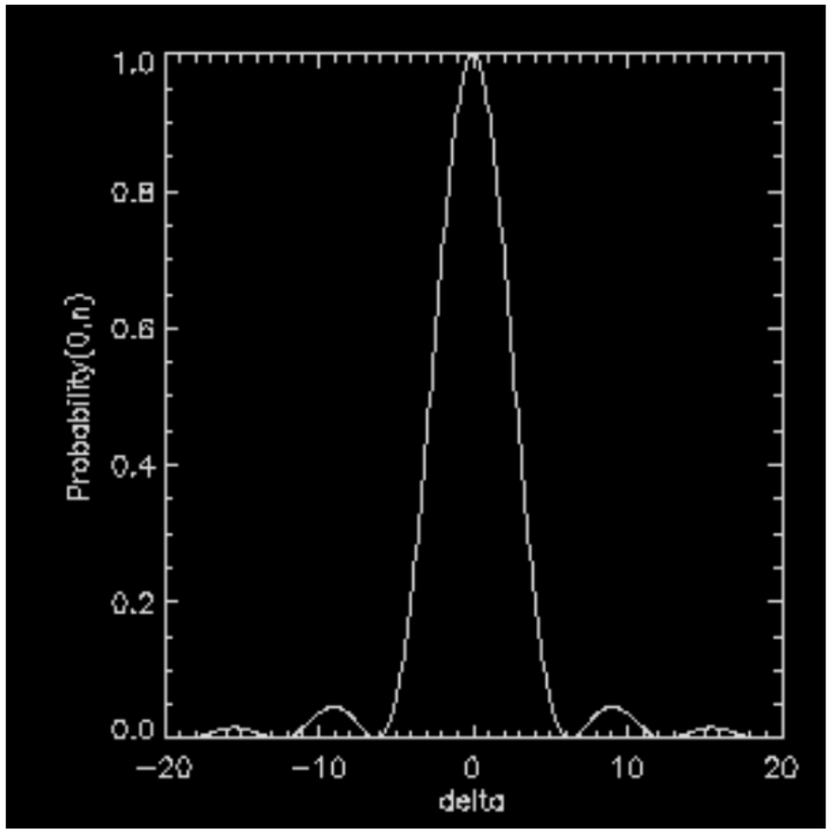
$$P_{0\to 1} = |\langle 1 \mid \psi(t)\rangle|^2 = \frac{q^2\mathcal{E}_0^2}{2m\hbar\omega}\left|\int_0^\infty dt' e^{i\omega t' - \frac{t}{\tau}}\right|^2$$

$$= \frac{q^2\mathcal{E}_0^2}{2m\hbar\omega}\frac{\tau^2}{\tau^2\omega^2 + 1}$$

We now return to the earlier general result we derived for the probability, namely,

$$P_{0\to n}(t) = \left|\langle n|\hat{V}_{t'}|0\rangle\right|^2\left(\frac{\sin\frac{\Delta t}{2\hbar}}{\Delta/2}\right)^2$$

In the Figure we plot this function.



Probability(0,n) versus Delta

The height of the central peak is proportional to t^2 and the location of the first zero is at

$$\Delta = \frac{2\pi\hbar}{t}$$

so that the width of the peak decreases as $1/t$.

The formula implies that for very short times

$$P_{0\to n} \propto t^2 \text{ for all } \varepsilon_n$$

As $t \to \infty$, however, the probability is largest for those states whose energy lies under the sharp bump near $\Delta = 0$ or those states with whose energy lies under the peak around ε_0.

Now the energy $\varepsilon_n \approx \varepsilon_0$ lies under the sharp bump when

$$|\Delta| = |\varepsilon_n - \varepsilon_0| < \frac{2\pi\hbar}{t}$$

The area under the bump is proportional to t and the rest of the area oscillates in time around zero.

This latter feature means that if $\varepsilon_n \neq \varepsilon_0$, the transition probability oscillates in time with a repetition time of

$$\frac{2\pi\hbar}{|\varepsilon_n - \varepsilon_0|}$$

The case, where we are looking for a transition to a single state, is, thus, only valid in perturbation theory for very small time t.

Otherwise the condition that the

$$P_{0\to 0}(t) \approx 1$$

will not be true and perturbation theory breaks down.

We also note that the probability cannot grow larger than one or that, after a while, the higher-order effects of the perturbation which we have neglected so far must become important and prevent the probability from exceeding one.

The condition that tells us whether a transition probability to a state with an energy appreciably different than the original energy is the same condition in time-independent perturbation theory that tells whether the state vector changes appreciably from the unperturbed state, namely

$$\left| \frac{\langle n | \hat{V} | 0 \rangle}{\varepsilon_n - \varepsilon_0} \right| << 1$$

Physically, a more interesting case occurs when the state $|n\rangle$ is one of a continuum of energy states, or it lies in a group of very closely spaced levels.

In this case we ask a different experimental question, namely,

> What is the probability that the system makes a transition to a small group of states near $|n\rangle$ (or has energy near ϵ_n)?

Since the area under the bump near $\Delta = 0$ or $\varepsilon_n \approx \varepsilon_0$ is proportional to t, we expect that the transition probability to a small group of states near ε_0 will grow linearly with t and thus

$$\frac{P_{0\to n}(t)}{t} = \text{ transition rate } = \Gamma = \text{ constant as } t \to \infty$$

Quantities that we measure are related to the transition rate and this result says that these measurements will make sense.

Let us now carry out this derivation in detail.

To calculate this transition rate we must sum $P_{0\to n}$ over the group of final states.

We assume that $|\langle b|\ \hat{V}_{t'}\ |0\rangle|^2$ is relatively constant over the small group of states near $|n\rangle$ (has a weak energy dependence).

We then have

$$\sum_{\substack{n \\ in group}} P_{0\to n}(t) = \left|\langle n|\,\hat{V}_{t'}\,|0\rangle\right|^2 \int_{group} d\varepsilon_n \rho(\varepsilon_n) \left[\frac{\sin\left[\frac{(\varepsilon_n-\varepsilon_0)t}{\hbar}\right]}{\left[\frac{(\varepsilon_n-\varepsilon_0)}{2}\right]}\right]^2$$

where

$\rho(\varepsilon_n)$ = number of states per unit energy

$\rho(\varepsilon_n)d\varepsilon_n$ = number of states in the interval $d\varepsilon_n$

Now in the limit $t\to\infty$

$$\left[\frac{\sin\left[\frac{(\varepsilon_n-\varepsilon_0)t}{\hbar}\right]}{\left[\frac{(\varepsilon_n-\varepsilon_0)}{2}\right]}\right]^2 \to \frac{2\pi t}{\hbar}\delta(\varepsilon_0-\varepsilon_n)$$

i.e., in general, for a sequence of functions

$$\delta_t(\alpha) = \frac{\sin^2 \alpha t}{\pi\alpha^2 t}$$

we have that

$$\delta_t(\alpha) = \begin{cases} \frac{t}{\pi} & \alpha = 0 \\ \leq \frac{1}{\pi\alpha^2 t} & \alpha \neq 0 \end{cases}$$

and

$$\lim_{t\to\infty} \int_{-\infty}^{\infty} d\alpha \delta_t(\alpha) F(\alpha) = F(0)$$

Therefore,

$$\lim_{t\to\infty} \delta_t(\alpha) = \delta(\alpha)$$

Using this result, we have

$$\sum_n P_{0\to n}(t) = \Gamma t$$

and thus

$$\Gamma = \text{ transition rate } = \frac{2\pi}{\hbar} \left|\langle n|\, \hat{V}\, |0\rangle\right|^2 \rho(\varepsilon_n)_{\varepsilon_n=\varepsilon_0}$$

which is called *Fermi's Golden Rule*.

We now consider a perturbation that depends explicitly on time.

In particular, suppose we have a harmonic perturbation of the form

$$\hat{V}_t = \hat{V} e^{\eta t} \cos\omega t = \frac{\hat{V}}{2} e^{\eta t} \left(e^{-i\omega t} + e^{i\omega t}\right)$$

and $|\psi(t_0)\rangle = |0\rangle$, where we let $t_0 \to -\infty$.

The $e^{\eta t}$ factor is necessary to make the mathematical operations valid in the limit.

It is equivalent for small η to turning the perturbation on slowly.

In the end we will let $\eta \to 0$.

We have

$$\langle n \mid \psi(t)\rangle = \langle n \mid 0\rangle + \frac{1}{i\hbar} e^{-\frac{i}{\hbar}\varepsilon_n t} \int_{-\infty}^{0} dt' \, \langle n|\hat{V}(t')\,|0\rangle$$

$$= \frac{e^{\eta t}}{2} \left[\frac{e^{i(\varepsilon_n - \varepsilon_0 - \hbar\omega)\frac{t}{\hbar}}}{\varepsilon_0 - \varepsilon_n + \hbar\omega + i\hbar\eta} + \frac{e^{i(\varepsilon_n - \varepsilon_0 + \hbar\omega)\frac{t}{\hbar}}}{\varepsilon_0 - \varepsilon_n - \hbar\omega + i\hbar\eta} \right] \langle n|\, \hat{V}\, |0\rangle$$

Thus, the probability of the energy being ε_n at the time t is

$$|\langle n \mid \psi(t)\rangle|^2$$
$$= \frac{e^{2\eta t}}{4}\left|\langle n|\hat{V}|0\rangle\right|^2 \left\{ \begin{array}{c} \frac{1}{(\varepsilon_0-\varepsilon_n+\hbar\omega)^2+(\hbar\eta)^2} + \frac{1}{(\varepsilon_0-\varepsilon_n-\hbar\omega)^2+(\hbar\eta)^2} \\ +2\mathrm{Re}\frac{e^{-2i\omega t}}{(\varepsilon_0-\varepsilon_n+\hbar\omega+i\hbar\eta)(\varepsilon_0-\varepsilon_n-\hbar\omega+i\hbar\eta)} \end{array} \right\}$$

The first term comes from the $e^{-i\omega t}$ part of $\hat{V}_t$ (positive frequency) and the second term comes from the $e^{i\omega t}$ part of $\hat{V}_t$ (negative frequency).

The last term represents *interference effects*.

Since

$$P_{0\to n}(t) = \Gamma_{0\to n} t$$

we have

$$\Gamma_{0\to n} = \frac{dP_{0\to n}(t)}{dt}$$

and thus

$$\Gamma_{0\to n}$$
$$= \frac{e^{2\eta t}}{4}\left|\langle n|\hat{V}|0\rangle\right|^2 \left\{ \begin{array}{c} \left[\frac{1}{(\varepsilon_0-\varepsilon_n+\hbar\omega)^2+(\hbar\eta)^2} + \frac{1}{(\varepsilon_0-\varepsilon_n-\hbar\omega)^2+(\hbar\eta)^2}\right](1-\cos 2\omega t) \\ +2\sin 2\omega t\left[\frac{e^{-2i\omega t}}{(\varepsilon_0-\varepsilon_n+\hbar\omega+i\hbar\eta)(\varepsilon_0-\varepsilon_n-\hbar\omega+i\hbar\eta)}\right] \end{array} \right\}$$

The sin ωt and cos ωt terms arise from the interference term.

In the limit $\eta \to 0$ and assuming that $|n\rangle$ is in the continuum part of spectrum, we have

1. the first two terms are not equal to zero only if $\varepsilon_n - \varepsilon_0 = \pm\hbar\omega$

2. the $\sin\omega t$ and $\cos\omega t$ terms average to zero if we assume that $\Gamma_{0\to n}$ is $dP_{0\to n}(t)/dt$ averaged over a few cycles of $\hat{V}_t$

which gives the result

$$\Gamma_{0\to n} = \frac{2\pi}{\hbar}\frac{\left|\langle n|\hat{V}|0\rangle\right|^2}{4}\left[\delta(\varepsilon_n - \varepsilon_0 - \hbar\omega) + \delta(\varepsilon_n - \varepsilon_0 + \hbar\omega)\right]$$

The positive and negative frequency parts act independently and the interference averages to zero.

Thus, the $e^{-i\omega t}$ part produced a $\Delta E > 0$ process (absorption), while the $e^{i\omega t}$ part produced a $\Delta E < 0$ process (emission).

To enhance our understanding of time-dependent perturbation theory, we look at a variation of this harmonic perturbation. Suppose we have a harmonic perturbation of the form

$$\hat{V}_t = \hat{V}e^{-i\omega t} + \hat{V}^+ e^{i\omega t}$$

which is only applied for a finite time interval $0 \le t \le T$.

If we start with energy ε_i in the state $|i\rangle$, then at any time $t \ge T$

$$a_f^{(1)}(t) = 1^{st} - \text{order amplitude for the state } |\psi(t)\rangle$$

$$\text{to have energy } \varepsilon_f \text{ (be in state } |f\rangle ?) \quad f \neq i$$

is given by

$$a_f^{(1)}(T) = \frac{1}{i\hbar}\langle f|\hat{V}|i\rangle\int_0^T e^{i(\omega_{fi}-\omega)t}dt + \frac{1}{i\hbar}\langle f|\hat{V}^+|i\rangle\int_0^T e^{i(\omega_{fi}+\omega)t}dt$$

and

$$\left|a_f^{(1)}(T)\right|^2 = \text{ probability that the final energy will be } \epsilon_f$$

As an example we consider spin resonance (we solved this problem exactly earlier).

We consider a spin = 1/2 particle in a static magnetic field B_0 (in the z–direction).

This says that the unperturbed Hamiltonian is

$$\hat{H}_0 = -\frac{1}{2}\hbar\gamma B_0 \hat{\sigma}_z$$

This operator has the eigenvectors and eigenvalues

$$|+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; |-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad , \quad \epsilon_\pm = \mp\frac{1}{2}\hbar\gamma B_0$$

We now perturb the system with another magnetic field B_1, which is rotating in the $x-y$ plane with angular velocity ω.

This implies that

$$\begin{aligned} \hat{V}_t &= -\frac{1}{2}\hbar\gamma B_1 (\cos\omega t \hat{i} + \sin\omega t \hat{j}) \cdot \hat{\sigma} \\ &= -\frac{1}{2}\hbar\gamma B_1 \left[\hat{\sigma}_x \cos\omega t + \hat{\sigma}_y \sin\omega t\right] = -\frac{1}{2}\hbar\gamma B_1 \begin{pmatrix} 0 & e^{-i\omega t} \\ e^{i\omega t} & 0 \end{pmatrix} \\ &= -\frac{1}{2}\hbar\gamma B_1 \left(\hat{V} e^{-i\omega t} + \hat{V}^+ e^{i\omega t}\right) \end{aligned}$$

where

$$\hat{V} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad , \quad \hat{V}^+ = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

We choose the initial state to be

$$|i\rangle = |n\rangle \;\; \text{(spin up in the z-direction)}$$

The first order perturbation theory approximation for this result is

$$\left|a_f^{(1)}(T)\right|^2 = \left(\frac{\omega_1}{\omega_0+\omega}\right)^2 \sin^2\frac{1}{2}(\omega_0+\omega)T$$

where

$$\omega_0 = \gamma B_0 \quad , \quad \omega_1 = \gamma B_1$$

and we have used

$$\langle -|\,\hat{V}\,|+\rangle = 0,\ \langle -|\,\hat{V}^+\,|+\rangle = -\frac{1}{2}\hbar\gamma B_1$$

$$\omega_{fi} = \frac{\varepsilon_f - \varepsilon_i}{\hbar} = \omega_0$$

When is the first order perturbation theory result valid?

If we compare the exact result with perturbation theory by expanding the exact result in a power series, we find that the two results agree exactly if

$$\left|\frac{\omega_1}{\omega_0+\omega}\right| << 1$$

which corresponds to $B_1 \ll B_0$ (as long as $\omega_0 + \omega \neq 0$).

When $\omega_0 + \omega = 0$ we have a phenomenon called resonance.

The exact solution gives

$$|a_f(T)|^2 = \sin^2\frac{1}{2}\alpha T$$

where $\alpha^2 = (\omega_0 + \omega)^2 + \omega_1{}^2$ and perturbation theory gives

$$\left|a_f^{(1)}(T)\right|^2 = \left(\frac{\omega_1 T}{2}\right)^2$$

Thus, the results agree only if $|\omega_1 T| \ll 1$ or if the perturbation only acts for a short time.

Atomic Radiation and Selection Rules

We now apply time-independent perturbation theory to the absorption and emission electromagnetic radiation by matter.

The Hamiltonian for an electron in an atom interacting with and electromagnetic field is

$$\hat{H} = \frac{\left(\vec{p}_{op} - \frac{q}{c}\vec{A}\right)^2}{2m_e} + q\phi + U$$

where $q = -e$, $U =$ the potential energy function that binds the electrons in the atom, and $\vec{A}$ and ϕ are the vector and scalar potentials associated with the electromagnetic field.

These potentials imply the electric and magnetic fields

$$\vec{E} = -\nabla\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t}, \vec{B} = \nabla \times \vec{A}$$

Note that if $\vec{A} = 0$, the $\vec{B} = 0$ and

$$\phi = -\int_0^{\vec{r}} \vec{E}(\vec{r}, t) \cdot d\vec{r}$$

We rewrite $\hat{H}$ as

$$\hat{H} = \hat{H}_0 + \hat{V}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{op}^2}{2m_e} + U \;=\; \text{Hamiltonian for the atom with no electromagnetic field}$$

and

$$\hat{V} = \frac{q}{2m_e c}\left(\vec{p}_{op} \cdot \vec{A} + \vec{A} \cdot \vec{p}_{op}\right) + \frac{q^2}{2m_e c^2}\left(\vec{A} \cdot \vec{A}\right) + q\phi$$

is the perturbation due to the presence of the electromagnetic field, i.e., the term $\hat{V}$ tells us how the atom interacts with the electromagnetic field.

The Electric Dipole Approximation

The typical wavelength of visible electromagnetic radiation is ≈ 5000 angstroms and the typical dimension of an atom is ≈ a few .

This implies that the electromagnetic fields are approximately constant over the volume of the atom.

In Gaussian units $|\vec{E}| \approx |\vec{B}|$, but the force due to $\vec{B} \approx (v/c) \times$ the force due to $\vec{E}$.

Thus, magnetic effects are negligible in most atoms compared to electric effects.

We therefore assume

1. $\vec{E} \approx$ constant over the volume of the atom
2. $\vec{B}$ can be neglected

This is the so-called electric dipole approximation.

In this approximation, we have A = 0 and as we said above, B = 0 and

$$\phi = -\int_0^{\vec{r}} \vec{E}(\vec{r}, t) \cdot d\vec{r}$$

This last integral is independent of path since

$$\nabla \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} = 0$$

Since we are assuming that $\vec{E} \approx$ constant over the volume of the atom, we get

$$\phi = -\vec{r} \cdot \vec{E}(t)$$

This adds a perturbation of the form

$$\hat{V} = -q\vec{r} \cdot \vec{E}(t)$$

to the Hamiltonian.

In some derivations the quantity

$$\hat{V} = \frac{q}{2m_e c}\left(\vec{p}_{op} \cdot \vec{A} + \vec{A} \cdot \vec{p}_{op}\right) + \frac{q^2}{2m_e c^2}\left(\vec{A} \cdot \vec{A}\right) + q\phi$$

is chosen as the perturbation.

The next step would be to expand in powers of the potentials.

Since $\nabla \cdot \vec{E} = 0$ for the radiation field, we can choose $\phi = 0$ and $\nabla \cdot \vec{A} = 0$ (if these relations were not true we could make them so with a gauge transformation).

Therefore, we end up with an expansion in powers of $\vec{A}$.

The first term in the expansion is

$$\hat{V}_{Ap} = -\frac{q}{2m_e c}\left(\vec{p}_{op} \cdot \vec{A} + \vec{A} \cdot \vec{p}_{op}\right) = -\frac{q}{m_e c}\vec{A} \cdot \vec{p}_{op}$$

since

$$\vec{p}_{op} \cdot \vec{A} - \vec{A} \cdot \vec{p}_{op} \propto \nabla \cdot \vec{A} = 0$$

In this case, in the electric dipole approximation

$$\vec{E} \approx \text{ constant} \quad , \quad \vec{B} = 0$$

$$\vec{A}(\vec{r}, t) = -c \int_0^{\vec{r}} \vec{E}(\vec{r}, t) dt$$

$$\phi(\vec{r}, t) = 0$$

If $\vec{E}(\vec{r},t) = \vec{E}(\vec{r})e^{-i\omega t}$, then $\vec{A} = e\vec{E}/i\omega$.

Using

$$\frac{\vec{p}_{op}}{m} = \frac{i}{\hbar}\left[\hat{H}_0, \vec{r}_{op}\right]$$

we get

$$\hat{V}_{Ap} = -\frac{q}{\hbar\omega}\left[\hat{H}_0, \vec{r}_{op}\right] \cdot \vec{E}$$

We can then calculate matrix elements

$$\begin{aligned}
\langle m|\, \hat{V}_{Ap}\, |n\rangle &= -\frac{q}{\hbar\omega}\, \langle m|\, \hat{H}_0 \vec{r}_{op} - \vec{r}_{op}\hat{H}_0\, |n\rangle \cdot \vec{E} \\
&= -\frac{q}{\hbar\omega}\, (\varepsilon_m - \varepsilon_n)\, \langle m|\, \vec{r}_{op} \cdot \vec{E}\, |n\rangle \\
&= \frac{\omega_{mn}}{\omega}\, \langle m|\, \hat{V}\, |n\rangle
\end{aligned}$$

where

$$\omega_{mn} = \frac{\varepsilon_m - \varepsilon_n}{\hbar}$$

Thus, the matrix elements of $\hat{V}_{A_p}$ (from $\vec{A}\cdot\vec{p}_{op}$) differ from the matrix elements of $\hat{V}$ (from ϕ) by the factor ω_{mn}/ω.

This implies different transition probabilities in first order except at resonance where $\omega_{mn}/\omega = 1$.

The reason for the differences is as follows:

1. we assumed that the perturbation = 0 for $t \le 0$, $t \ge T$
2. there is no physical problem with $V = -q\vec{r} \cdot \vec{E}$ changing discontinuously
3. however, if $\vec{A} \cdot \vec{p}_{op}$ changes discontinuously, then the relation

$$\vec{E} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial t}$$

generates spurious δ–function type $\vec{E}$ fields

It is clear that one must exercise great care in choosing a starting point for perturbation theory.

Induced Emission and Absorption

We now look at the physics connected with the harmonic perturbation.

Physically, it represents an electromagnetic wave interacting with the atom.

We consider the perturbing potential

$$\hat{V} = \begin{cases} -q\vec{r} \cdot \vec{E}_0 (e^{-i\omega t} + e^{i\omega t}) & 0 < t < T \\ 0 & t < 0, t > T \end{cases}$$

where $\vec{E}_0$ = a constant vector which tells us the strength and polarization of the electromagnetic field.

This perturbing potential corresponds to monochromatic(single wavelength) electromagnetic radiation.

For an initial state $|\psi(0)\rangle = |i\rangle$ where $\hat{H}_0 |i\rangle = \varepsilon_i |i\rangle$ the probability, at any time $t \ge T$, that the atom will have a final energy ε_f is

$$P_{i \to f}(T) = \left| a_f^{(1)}(T) \right|^2$$

to first order, where we have from our earlier derivation

$$a_f^{(1)}(T) = \frac{\langle f|\left(-q\vec{r}\cdot\vec{E}_0\right)|i\rangle}{\hbar}\left[\frac{1-e^{i(\omega_{fi}-\omega)T}}{\omega_{fi}-\omega}+\frac{1-e^{i(\omega_{fi}+\omega)T}}{\omega_{fi}+\omega}\right]$$

If $\varepsilon_f > \varepsilon_i$ this gives the probability amplitude for absorbing radiation and if $\varepsilon_f < \varepsilon_i$ this gives the probability amplitude for emitting radiation.

In the limit $T \to \infty$, as we saw earlier, $P_{i\to f}(t) = \Gamma_{i\to f}\, t$ which gives

$$\Gamma_{i\to f} = \frac{2\pi q^2}{\hbar}\frac{\left|\langle f|\,\vec{r}\cdot\vec{E}_0\,|i\rangle\right|^2}{4}\left[\delta(\varepsilon_i-\varepsilon_f-\hbar\omega)+\delta(\varepsilon_i-\varepsilon_f+\hbar\omega)\right]$$

In this expression

$$\delta(\varepsilon_i-\varepsilon_f-\hbar\omega) \to \text{ absorption}$$
$$\delta(\varepsilon_i-\varepsilon_f+\hbar\omega) \to \text{ emission}$$

We then have the transition rates

$$\Gamma^a_{i\to f} = \Gamma^e_{i\to f} = \frac{2\pi q^2}{\hbar}\frac{\left|\langle f|\,\vec{r}\cdot\vec{E}_0\,|i\rangle\right|^2}{4}$$

These expressions are *zero/nonzero* or transitions are *not-allowed/allowed* depending on the matrix element of the perturbation, i.e.,

These relationships between the quantum numbers of the initial and final states that tell us whether or not a transition is allowed are called *selection rules*.

To determine the selection rules for one-electron atoms we only need to consider matrix elements of the form

$$\langle n'\ell' m_{\ell'} m_{s'} | \, \vec{r} \cdot \vec{E}_0 \, | n\ell m_\ell m_s \rangle$$

or we need to look at three matrix elements, namely,

$$\langle n'\ell' m_{\ell'} m_{s'} | \, \hat{x} \, | n\ell m_\ell m_s \rangle \, , \langle n'\ell' m_{\ell'} m_{s'} | \, \hat{y} \, | n\ell m_\ell m_s \rangle \text{ and } \langle n'\ell' m_{\ell'} m_{s'} | \, \hat{z} \, | n\ell m_\ell m_s \rangle$$

Now

$$z = r\cos\theta \propto rY_{10} = x_3$$
$$x = r\sin\theta\cos\varphi \propto r(Y_{11} + Y_{1,-1}) = x_1$$
$$y = r\sin\theta\sin\varphi \propto r(Y_{11} - Y_{1,-1}) = x_2$$

A typical matrix element, therefore, will have a term like the following:

$$\langle sm_{s'} \mid sm_s \rangle \int R_{n'\ell'}(r) R_{n\ell}(r) r^3 dr \int Y^*_{\ell' m\ell'} Y_{\ell m_\ell} Y_{1m} d\Omega$$

where $m = \pm 1, 0$.

The radial integral would equal zero only by accident implying that it is not part of the general selection rules, which must come from the other terms.

The term $\langle sm_{s'} \mid sm_s \rangle = \delta_{ms'\ ms}$ gives us a simple selection rule (this simple rule arises here because the interaction does not depend on spin).

We have

$$m_{s'} = m_s \rightarrow \text{ the } \Delta m_s = 0 \qquad \textbf{SELECTION RULE}$$

The rest of the selection rules come from the angular integration terms

$$\int Y^*_{\ell' m_{\ell'}} Y_{\ell m_\ell} Y_{1m} d\Omega$$

This integral equals 0 unless $\ell + \ell' + 1 =$ an even integer.

This rule follows from parity considerations.

For any angular integration over all angles to be nonzero, the integrand must be even under the parity operation.

Now

$$Y_{\ell m} \to (-1)^{\ell} Y_{\ell m}$$

under the parity operation.

Therefore

$$Y^*_{\ell' m \ell'} Y_{\ell m_\ell} Y_{1m} \to (-1)^{\ell+\ell'+1} Y^*_{\ell' m \ell'} Y_{\ell m_\ell} Y_{1m}$$

which gives the stated rule.

The θ–integration says that we must have

$$|\ell' - \ell| \leq 1 \leq \ell' + \ell$$

This corresponds to thinking of the integrand as made up of two states, namely,

$$\langle \ell' m_{\ell'} | \left(|\ell m_\ell\rangle \otimes |1m\rangle \right)$$

Our angular momentum addition rules say that

$$|\ell m_\ell\rangle \otimes |1m\rangle = |\ell+1\rangle \oplus |\ell\rangle \oplus |\ell-1\rangle$$

and the selection rule then follows from the orthogonality condition.

The two ℓ –rules when combined imply the selection rule

$$\Delta m_\ell = m_{\ell'} - m_\ell = \pm 1, 0 \qquad \textbf{SELECTION RULE}$$

Thus, for transitions within the electric dipole approximation, as defined above, we have the SELECTION RULES

$$\Delta m_\ell = m_{\ell'} - m_\ell = \pm 1\,,\, 0$$
$$\Delta \ell = \ell' - \ell = \pm 1$$
$$\Delta m_s = 0 = \Delta s$$

The derivation is more complex for multi-electron atoms due the complexity of the wave function, but it can be shown that, in general, the SELECTION RULES in the electric dipole approximation are

1. parity changes
2. $\Delta(\sum \ell_i) = \pm 1$
3. $\Delta S = 0$
4. $\Delta M_S = 0$
5. $\Delta L = \pm 1, 0$ (possibility of 0 which is not allowed for 1-electron atoms)
6. $\Delta M_L = \pm 1, 0$
7. $\Delta J = \pm 1, 0$
8. $\Delta M_J = \pm 1, 0$
9. $J = 0 \rightarrow J = 0$ is strictly forbidden

A Real Physical Process - Ionization

We now carry out the calculations of the transition rate for the ionization(transition to the continuum) of hydrogen by electromagnetic radiation.

The initial state of the electron is the ground state of hydrogen

$$|i\rangle = |100\rangle \text{ with energy } \varepsilon_i = \varepsilon_{100} = -\frac{e^2}{2a_0}$$

The final state of the electron in the ionization process is a free particle state(ionized electron)

$$|f\rangle = \left|\vec{k}\right\rangle \text{ with energy } \varepsilon_f = \varepsilon_k = \frac{\hbar^2 k^2}{2m}$$

We then have

$$\vec{p}_{op}\left|\vec{k}\right\rangle = \hbar\vec{k}\left|\vec{k}\right\rangle$$

which says this is a momentum eigenstate also (for free particles $[\hat{H}, \vec{p}_{op}] = 0$ and momentum and energy have the same eigenstates).

The momentum is given by $\vec{p} = \hbar\vec{k}$.

Since $\varepsilon_f > \varepsilon_i$ this is an absorption process and thus the transition rate is given by

$$\Gamma_{0\to\vec{k}} = \text{transitionrate} = \lim_{t\to\infty} \frac{P_{0\to\vec{k}}(t)}{t}$$

$$= \frac{2\pi}{\hbar}\left|\left\langle\vec{k}\right|\hat{V}\left|100\right\rangle\right|^2 \delta(\varepsilon_k - \varepsilon_{100} - \hbar\omega)$$

where ω = frequency of the electromagnetic radiation and

$\hat{V} = -e\vec{r}\cdot\vec{\mathcal{E}}$ (electric dipole approximation)

$\vec{\mathcal{E}}$ = the electric field vector

We define

$$d\Gamma = \text{ rate of transition into a small solid angle } d\Omega$$
$$= \sum_{\vec{k} \text{ in } d\Omega} \Gamma_{0\to\vec{k}}$$

For convenience we use a common trick and assume that the universe is a large box (side = L, volume = L^3).

This allows us to normalize the plane wave states associated with the free electron.

We have

$$\langle \vec{r} \mid \vec{k} \rangle = Ae^{i\vec{k}\cdot\vec{r}}$$
$$\langle \vec{k} \mid \vec{k} \rangle = 1 = \int d^3\vec{r} \langle \vec{k} \mid \vec{r} \rangle \langle \vec{r} \mid \vec{k} \rangle$$
$$= A^2 \int d^3\vec{r} e^{-i\vec{k}\cdot\vec{r}} e^{i\vec{k}\cdot\vec{r}} = A^2 \int d^3\vec{r} = A^2 L^3$$

or

$$A = \frac{1}{L^{3/2}}$$

Now there are

$$L^3 \frac{d^3\vec{k}}{(2\pi)^3} = \frac{L^3}{(2\pi)^3} d\Omega k^2 dk = \frac{L^3 mk}{(2\pi)^3 \hbar^2} d\Omega d\varepsilon_k$$

states in the volume d^3k of phase space.

This implies that there are

$$\frac{L^3 mk}{(2\pi)^3 \hbar^2}$$

states per unit energy per unit solid angle.

Therefore

$$d\Gamma = \sum_{\vec{k}\text{in}d\Omega} \Gamma_{0\to\vec{k}} \to d\Omega \int_0^\infty \frac{L^3 mk}{(2\pi)^3\hbar^2} d\varepsilon_k \frac{\pi e^2}{2\hbar} \left|\left\langle \vec{k}\right| \vec{r}\cdot\vec{\mathcal{E}} \left|100\right\rangle\right|^2 \delta(\varepsilon_k - \varepsilon_{100} - \hbar\omega)$$

Doing the integration (using the delta function) we get

$$d\Gamma = d\Omega \frac{L^3 mke^2}{16\pi^2\hbar^3} \left|\left\langle \vec{k}\right| \vec{r}\cdot\vec{\mathcal{E}} \left|100\right\rangle\right|^2$$

where

$$\varepsilon_k = \varepsilon_{100} + \hbar\omega = \frac{\hbar^2 k^2}{2m}$$

$$k = \left(\frac{2m\omega}{\hbar} - \frac{1}{a_0^2}\right)^{1/2}$$

Evaluation of the Matrix Element

We have

$$\begin{aligned}
\left\langle \vec{k}\right| \vec{r}\cdot\vec{\mathcal{E}} \left|100\right\rangle &= \int d^3\vec{r} \left\langle \vec{k} \mid \vec{r}\right\rangle \left\langle \vec{r}\right| \vec{r}\cdot\vec{\mathcal{E}} \left|100\right\rangle \\
&= \frac{1}{L^{3/2}} \int d^3\vec{r} e^{-i\vec{k}\cdot\vec{r}} \vec{r}\cdot\vec{\mathcal{E}} \left\langle \vec{r} \mid 100\right\rangle \\
&= \frac{1}{L^{3/2}} \int d^3\vec{r} e^{-i\vec{k}\cdot\vec{r}} \vec{r}\cdot\vec{\mathcal{E}} R_{10}(r) Y_{00} \\
&= \frac{2}{a_0^{3/2} L^{3/2}} \int d^3\vec{r} e^{-i\vec{k}\cdot\vec{r}} e^{-\frac{r}{a_0}} \vec{r}\cdot\vec{\mathcal{E}}
\end{aligned}$$

We can arbitrarily choose $\vec{k} = k\hat{z}$ which gives

$$\left\langle \vec{k}\right| \vec{r}\cdot\vec{\mathcal{E}} \left|100\right\rangle = \frac{2}{a_0^{3/2} L^{3/2}} \int d\Omega dr r^2 e^{-ikr\cos\theta - \frac{r}{a_0}} \vec{r}\cdot\vec{\mathcal{E}}$$

using $\vec{\mathrm{k}}\cdot\vec{\mathrm{R}} = kz = kr\cos\sigma$.

Now in spherical-polar coordinates we have

$$\vec{\mathcal{E}} = \mathcal{E}\left(\sin\theta_\varepsilon \cos\varphi_\varepsilon \hat{e}_x + \sin\theta_\varepsilon \sin\varphi_\varepsilon \hat{e}_z + \cos\theta_\varepsilon \hat{e}_z\right)$$
$$\vec{r} = r\left(\sin\theta\cos\varphi\hat{e}_x + \sin\theta\sin\varphi\hat{e}_y + \cos\theta\hat{e}_z\right)$$

so that

$$\vec{\mathcal{E}}\cdot\vec{r} = \mathcal{E}r\left(\cos\theta\cos\theta_\varepsilon + \sin\theta\sin\theta_\varepsilon\cos(\varphi-\varphi_\varepsilon)\right)$$

We then have

$$\left\langle\vec{k}\middle|\vec{r}\cdot\vec{\mathcal{E}}\middle|100\right\rangle$$
$$= \frac{2\varepsilon}{a_0^{3/2}L^{3/2}}\int_0^{2\pi} d\varphi \int_0^{\pi}\sin\theta d\theta \int dr r^3 e^{-ikr\cos\theta-\frac{r}{a_0}}\left[\cos\theta\cos\theta_\varepsilon + \sin\theta\sin\theta_\varepsilon\cos(\varphi-\varphi_\varepsilon)\right]$$

Since

$$\int_0^{2\pi} d\varphi\cos\varphi = \int_0^{2\pi} d\varphi\sin\varphi = 0$$

the $\phi-$ integration wipes out the $\cos(\phi-\phi_\varepsilon)$ term.

Letting $x = \cos\theta$ we then have

$$\left\langle\vec{k}\middle|\vec{r}\cdot\vec{\mathcal{E}}\middle|100\right\rangle = \frac{4\pi\mathcal{E}\cos\theta_\varepsilon}{a_0^{3/2}L^{3/2}}\int_{-1}^{1}\left[\int_0^{\infty} dr r^3 e^{-ikrx-\frac{r}{a_0}}\right]x dx$$

Now

$$\int_0^{\infty} dr r^3 e^{-ikrx-\frac{r}{a_0}} = \int_0^{\infty} dr r^3 e^{-\alpha r} = \frac{3!}{\alpha^4} = \frac{6}{\left(ikx+\frac{1}{a_0}\right)^4}$$

Therefore, we have

$$\langle \vec{k} | \vec{r} \cdot \vec{\mathcal{E}} | 100 \rangle = \frac{4\pi \mathcal{E} \cos\theta_\varepsilon}{a_0^{3/2} L^{3/2}} \int_{-1}^{1} \frac{6x}{\left(ikx + \frac{1}{a_0}\right)^4} dx$$

$$= \frac{4\pi \mathcal{E} \cos\theta_\varepsilon}{a_0^{3/2} L^{3/2}} \frac{16 k a_0^5}{(1 + k^2 a_0^2)^3}$$

and

$$\left| \langle \vec{k} | \vec{r} \cdot \vec{\mathcal{E}} | 100 \rangle \right|^2 = \frac{4096 \pi^2 \mathcal{E}^2 \cos^2\theta_\varepsilon a_0^7 k^2}{L^3 (1 + k^2 a_0^2)^6}$$

where

$$k^2 = \frac{2m\omega}{\hbar} - \frac{1}{a_0^2}$$

Finally, we get

$$\Gamma_{ionization} = \int d\Gamma = \frac{L^3 m k e^2}{16\pi^2 \hbar^3} \int d\Omega_{\vec{k}} \left| \langle \vec{k} | \vec{r} \cdot \vec{\mathcal{E}} | 100 \rangle \right|^2$$

where

$$d\Omega_{\vec{k}} = \text{ integration over the angles of } \vec{k}$$
$$\text{(varies direction of arbitrary z - axis)}$$

However, varying the z-direction is the same as keeping $\vec{\mathrm{E}}$ fixed and integrating over $d\Omega_{\vec{\mathrm{E}}}$.

Therefore, we have

$$\Gamma_{ionization} = \frac{L^3 m k e^2}{16\pi^2 \hbar^3} \frac{4096 \pi^2 \mathcal{E}^2 a_0^7 k^2}{L^3 (1 + k^2 a_0^2)^6} \int d\Omega_{\vec{\mathcal{E}}} \cos^2\theta_\varepsilon$$

$$= \frac{64\pi m e^2 \mathcal{E}^2 a_0^7}{3\hbar^3} \frac{k^3}{(1 + k^2 a_0^2)^6}$$

Now

$$1 + k^2 a_0^2 = \frac{2m\omega a_0^2}{\hbar}$$

and letting

$$\omega_0 = \frac{\hbar}{2ma_0^2}$$

we find that

$$1 + k^2 a_0^2 = \frac{\omega}{\omega_0} \text{ and } k = \frac{1}{a_0}\left(\frac{\omega}{\omega_0} - 1\right)^{1/2}$$

and we get

$$\Gamma_{ionization} = \frac{64\pi e^2 \mathcal{E}^2 a_0^3}{3\hbar}\left(\frac{\omega_0}{\omega}\right)^6 \left(\frac{\omega}{\omega_0} - 1\right)^{3/2}$$

Thus, there exists a threshold energy for this process, i.e., it cannot occur unless the energy of the photon is greater than a minimum amount, which makes physical sense.

In particular, we must have $\omega \geq \omega_0$ so that

$$\hbar\omega_0 = |\varepsilon_{100}| = E_{ionization} = \text{ minimum} = \text{threshold}$$

We also get the correct 6^{th}–power term in the answer which agrees with experiment.

The remainder of these notes is a presentation of the

Quantum Measurement notes (Chapter 10 in book)

from the Agnes Scott Lifelong Learning class.

You now understand

so much more fundamental Quantum Mechanics

that you can extract

a much greater understanding

of the ideas and concepts presented.

Basic Quantum Mechanics Reviewed

We now review the standard formulation of quantum mechanics that we have developed,

which is based on the 4 postulates listed below(with some embellishments):

- All physical systems are represented by ket vectors $|\psi\rangle$ normalized to 1, i.e., $\langle\psi|\psi\rangle = 1$. The ket labels represent **everything that we know about the system**.

- Measurable properties of physical systems are represented by linear operators called **observables**.

So restating part of the first postulate, the ket labels represent **the values of all observables of the system that have been measured - we PREPARED the state!!!**

If a vector associated with a particular state $|\psi\rangle$ is an eigenvector, with eigenvalue α, of operator $\hat{A}$ associated with a particular measurable property of the system, i.e., if $\hat{A}\,|\psi\rangle = \alpha\,|\psi\rangle$, then the system in that state **definitely** has the value α of that measurable property.

This implies that if one performs a measurement corresponding to the observable represented by A on a system in the state $|\psi\rangle$, then with **certainty** (probability = 1) the measurement yields the value α for that measurable property.

Observables are represented by **Hermitian** operators (real eigenvalues). Since the eigenvectors of any Hermitian operator form a complete, orthonormal set, they can be used as a **basis** for the Hilbert space of the system

Finally, if the system is in the state |ψ⟩ and one measures an observable $\hat{B}$, where |ψ⟩ is **not** an eigenvector of $\hat{B}$, then the **only possible results** of the measurement are one of the eigenvalues $\{b_k\}$ of $\hat{B}$.

- Dynamics of state vectors

  The state vectors of any system change with time via **deterministic** laws (similar to classical rules).

  We define the **time evolution or time development operator** that governs how a state vector changes in time by the relationship

$$|A, t + \Delta t\rangle = \hat{U}(\Delta t)\,|A, t\rangle$$

  or the state vector at time t+ Δt is given by the time evolution operator (Unitary) $\hat{U}$ operating on the state vector at time t.

  In general, the ket labels(which contain whatever we know (have measured) about the state) are the only thing that changes.

The time evolution operator is a unitary operator since the state vector must remain normalized to 1, i.e., the vector length cannot change, and this is guaranteed by the use of a unitary operator. The only changes to state vectors in quantum mechanics are changes in direction (phase).

The time evolution operator is related to the energy operator (this follows from time translation invariance)

$$\hat{U}(t) = e^{-i\hat{H}t/\hbar}$$

- **Connection with Experiment/Measurements**

We have specified above what happens when one measures a certain property of a physical system at the moment when the state vector of the system is an eigenvector of the operator representing the property - probability = 1 that we get the corresponding eigenvalue.

What if one measures a certain property of a physical system at a moment when the state vector of the system does not happen to be an eigenvector of the operator representing that property (which is most of the time) - what are the probabilities?

We needed a new assumption.

Suppose the system is in the state $|\psi\rangle$, and one carries out a measurement of a property (observable) associated with the operator $\hat{B}$. We assume the eigenvectors of $\hat{B}$ are the vectors (states) $|b_i\rangle$, which means that $\hat{B}\,|b_i\rangle = b_i\,|b_i\rangle$, i=1,2,3,…. where the b_i are the corresponding eigenvalues.

Quantum theory now assumes that the outcome of measurement is strictly a matter of probability.

Quantum theory stipulates that the probability that the outcome of a measurement of $\hat{B}$ on the state $|\psi\rangle$ (not an eigenvector) will yield the result b_i (remember the only possible results of measurement are the eigenvalues of $\hat{B}$ no matter what state the system is in), is equal to $|\langle b_i\,|\,\psi\rangle|^2$ (the **Born rule**).

The probability is given by the absolute square of the corresponding component.

remember that — — component = amplitude

The quantum mechanics formalism based on these postulates + embellishments correctly predicts experimental results for all known experiments (over 100 years).

Some ideas implied by these rules are:

These rules imply that one **cannot** say anything definite about the value of the observable represented by $\hat{B}$ when system is in a state $|\psi\rangle$, which is NOT an eigenvector of $\hat{B}$.

One can only make probability statements.

Before one measures the observable represented by $\hat{B}$ when the system is in a state $|\psi\rangle$, which is NOT an eigenvector of $\hat{B}$, the system **does not** have a value of that observable, according to quantum theory!

Our information about any state is only a set of probabilities.

But all of your experience says that objects have values for measured quantities before they are measured, i.e., your experience tells you that the observable represented by $\hat{B}$ has a value even if we do not measure it.

That is your view (the standard classical view) about what is real and what is not real.

Quantum theory implies you are wrong in both cases!!

Where is the "collapse" postulate?

Since the system has a definite value of the observable represented by $\hat{B}$ after the measurement, i.e., a pointer points to a value or a counter clicks or a mark is registered on a piece of paper (note that these are all **irreversible occurrences** which must be the end result of any measurement) and there is no mechanism to produce a single value in the rules as presented so far, how does it happen?

Most presentations add another rule(as we did earlier) at this point called **collapse of the state vector.**

It is usually proposed that the effect of a measurement is to irreversibly change(collapse) the state vector (which was not an eigenvector of $\hat{B}$) into an eigenvector of $\hat{B}$ (corresponding to the eigenvalue just measured) so that it would be observed to have a definite (probability = 1) value for a subsequent measurement of the operator $\hat{B}$.

This extra rule says that state vector changes(discontinuously) during measurement from representing range of possibilities (superposition of all possible states) to definite state or only one possible outcome.

Which particular eigenvector it gets changed into is determined by outcome of measurement and cannot be known until then!.

It cannot be predicted! It is at this point that randomness enters quantum mechanics.

I believe that this last rule is not needed and should not be added.

First of all, no real mechanism is ever given for "how" this process actually takes place, and second no specifications are given as to exactly "when" it occurs.

Also, we don't need collapse mechanism to make any predictions confirmed by experiments.

That suggests to me that it does not exist! That it is NOT a real process!!

I will now proceed to develop a proposal for "definite outcomes" **without** using any "collapse" rule.

The path I will follow will not be direct, but it will represent the actual tortuous path I followed to get to the conclusions.

I will meander about, repeating myself often on purpose so I can make subtle changes to the previous ideas as I learn new wrinkles along the way.

I will initially end up with one conclusion, which we will see has flaws and then I will fix the flaws and get what I think is the correct answer at the end.

The Measurement Process(some repetition of earlier discussions)

We consider a system consisting of a quantum system (Q-system) and a measurement system (M-system).

If the meter, which we assume is initially OFF (state $|0\rangle_M$) was turned ON when quantum system was in $|+\rangle_Q$ state, then according to the above rules the combined system evolves to

$$|+\rangle_Q |0\rangle_M \rightarrow |+\rangle_Q |+1\rangle_M$$

i.e., meter (assuming a good meter) reads +1.

Similarly, if the meter turned ON when the system is in the $|-\rangle_Q$ state, then combined system evolves to

$$|-\rangle_Q |0\rangle_M \rightarrow |-\rangle_Q |-1\rangle_M$$ i.e., meter (assuming a good meter) reads –1.

This indicates that measurement, within framework of our rules, CORRELATES or ENTANGLES the dynamical variables (Q-system) being measured and the macroscopic (M-system) indicator of the meter, which we assume can be directly (macroscopically) observed (and is **irreversible**).

Let us expand(and repeat) this discussion a bit.

We have supposed above that the meter has eigenvectors (labelled by the corresponding eigenvalues), i.e., it is a quantum system also.

$$\begin{aligned} |+\rangle_M &\Rightarrow meter \text{ on}: \ reading \ +1 \\ |-\rangle_M &\Rightarrow meter \text{ on}: \ reading \ -1 \\ |0\rangle_M &\Rightarrow meter \text{ off} \end{aligned}$$

and the quantum system has eigenvectors (labelled by eigenvalues)

$$\begin{aligned} |+\rangle_Q &\Rightarrow value = +1 \\ |-\rangle_Q &\Rightarrow value = -1 \end{aligned}$$

Now suppose that the initial state of the quantum system is a **superposition**

$$|\psi\rangle = a\,|+\rangle_Q + b\,|-\rangle_Q$$

and thus the initial state of the combined system is given by

$$|initial\rangle = \left(a\,|+\rangle_Q + b\,|-\rangle_Q\right)|0\rangle_M$$

which represents the system in a superposition state and meter off.

We are interested in the evolution of this state according to QM.

We note as stated above, if, instead of the above initial state, we started with initial states

$$|A\rangle = |+\rangle_Q\,|0\rangle_M \qquad \textbf{OR} \qquad |B\rangle = |-\rangle_Q\,|0\rangle_M$$

and then turn on the meter, these states must evolve as

$$|A\rangle = |+\rangle_Q\,|0\rangle_M \longrightarrow |A'\rangle = |+\rangle_Q\,|+\rangle_M$$

$$|B\rangle = |-\rangle_Q\,|0\rangle_M \longrightarrow |B'\rangle = |-\rangle_Q\,|-\rangle_M$$

respectively, indicating that the meter measured the appropriate value (definition of good meter) since system is in eigenstate and has a definite value with certainty.

If system is in initial state corresponding to a superposition, however, then the **linearity** of quantum mechanics says must it evolve into

$$|initial\rangle = \left(a\,|+\rangle_Q + b\,|-\rangle_Q\right)|0\rangle_M \longrightarrow |final\rangle = a\,|+\rangle_Q\,|+1\rangle_M + b\,|-\rangle_Q\,|-1\rangle_M$$

We note a problem immediately,

i.e., the meter has not ended up in a state with a definite value

- it remains in a superposition of two macroscopically different pointer readings,

which is **never** observed in the real world.

Hence, if as most physicists assume,

the state vector represents a complete description of the Q-system,

there seems to be a need for the **"collapse"** rule

to fix the result and obtain **"definite"** values!

Since we **will not** be incorporating the "collapse" rule, we must proceed in a **different** way.

We **cannot use** the state vector as the fundamental object in quantum theory

since it **seems** that it must collapse (or maybe not!) for quantum mechanics to work

and we are **not** including "collapse".

Introduce new mathematical object called the Density Operator to remedy this situation:

Now we present this **alternate** way of representing quantum states. Reasons why come later!

It is still an open question whether this alternative is a **simple** mathematical convenience or a more ontologically **true** representation.

Either way, it has a key role to play in modern interpretations of quantum theory and also leads to a possible solution to measurement problem.

This discussion is more mathematical than rest of these notes, but you will benefit if you persevere and work through material. **It is very important**.

The problem, as we will now see, lies with assuming that the state vector is the proper way to represent the Q-system **during** the measurement process.

Now, the expectation value is the **average** of set of measurement results taken from a collection of systems in **same** state.

A straightforward calculation of the expectation value in a specific state takes the following form, with $\hat{O}$ being an operator representing the measurement of the specific physical variable and $|\phi\rangle$ is a state vector of some system in the collection:

$$\left\langle \hat{O} \right\rangle = \langle \phi | \, \hat{O} \, | \phi \rangle$$

If we choose **any set** of basis vectors {|i⟩}, i = 1, 2, ... for our vector space, we can expand |ϕ⟩ and ⟨ϕ| as

$$|\phi\rangle = \sum_i a_i |i\rangle \quad , \quad \langle\phi| = \sum_j a_j^* \langle j|$$

where

$$a_i = \langle i | \phi \rangle \quad , \quad a_j^* = \langle \phi | j \rangle$$

Plugging these expansions into the expression for the expectation value:

substitution

rearrangement

$$\langle \hat{O} \rangle = \sum_j a_j^* \langle j| \hat{O} \sum_i a_i |i\rangle = \sum_j \left(\sum_i a_j^* a_i \langle j| \hat{O} |i\rangle \right)$$

$$= \sum_j \left(\sum_i \langle \phi | j \rangle \langle i | \phi \rangle \langle j| \hat{O} |i\rangle \right) = \sum_j \left(\sum_i \left[\langle i | \phi \rangle \langle \phi | j \rangle \langle j| \hat{O} |i\rangle \right] \right)$$

more rearrangement

more rearrangement

Now we define a new operator $\hat{\rho}$ = |ϕ⟩⟨ϕ|, which, at this point, is just the **projection operator** onto the state |ϕ⟩.

We can then write the expectation value as

$$\langle \hat{O} \rangle = \sum_i \left(\sum_j \langle i| \hat{\rho} |j\rangle \langle j| \hat{O} |i\rangle \right)$$

substitution of new definition

Recognizing the relation $\sum_j |j\rangle\langle j| = \hat{I}$ appearing in the expression

one finds that the expectation value can be rewritten in a very interesting form

$$\langle \hat{O} \rangle = \sum_i \left(\sum_j \langle i| \hat{\rho} |j\rangle \langle j| \hat{O} |i\rangle \right) = \sum_i \left(\langle i| \hat{\rho} \left(\sum_j |j\rangle \langle j| \right) \hat{O} |i\rangle \right)$$

$$\langle \hat{O} \rangle = \sum_i \langle i| \hat{\rho} \hat{I} \hat{O} |i\rangle = \sum_i \langle i| \hat{\rho} \hat{O} |i\rangle = \mathrm{Tr}(\hat{\rho}\hat{O})$$

<— the important result!!

sum over diagonal matrix elements

i.e., the expectation value is given by the sum over the **diagonal** matrix elements of the operator product $\hat{\rho}\hat{O}$ (the symbol Tr = Trace is just shorthand for the diagonal sum).

The new operator $\hat{\rho}$ is called the **density operator**.

Why bother to introduce this new operator?

As we will see, the **real power** of density operator approach to QM comes when we have to deal with a situation in which we cannot be sure what state the system is in (just as is the case in the measurement problem!!).

Now, imagine we have a whole collection of identical systems, some in $|\phi_1\rangle$, some in $|\phi_2\rangle$, etc.

We might not know which system in which state, and might not even know how many systems are in any one state.

Example: Think about a beam of electrons that **has not** passed through any Stern-Gerlach (S-G) magnet.

Chances are that the spin states of the electrons are completely random.

Perhaps the best one can know is the probability of finding an electron in any state.

$$P_1 = \mathrm{Prob}(|\phi_1\rangle) \quad , \quad P_2 = \mathrm{Prob}(|\phi_2\rangle) \quad , \quad P_3 = \mathrm{Prob}(|\phi_3\rangle) \quad , \quad .....$$

These probabilities have **nothing to do** with quantum theory.

They simply represent our **ignorance** of the details of what is happening in the beam.

Thus, they are **not related** to any quantum amplitudes.

Given a situation like this, one should still be able to do some useful calculations.

For example, one could work out the expectation value of any measurement as follows.

If one can calculate the expectation value of each individual state, then the overall expectation value is simply given by

$$\left\langle \hat{O} \right\rangle = P_1 \langle \phi_1 | \hat{O} | \phi_1 \rangle + P_2 \langle \phi_2 | \hat{O} | \phi_2 \rangle + P_3 \langle \phi_3 | \hat{O} | \phi_3 \rangle + ... + P_n \langle \phi_n | \hat{O} | \phi_n \rangle$$

that is just the standard definition of the average value!!

Think back to original definition of expectation value which just represents the average value of measurement and this will be clear.

What we have done here is put together the weighted average of the average value for each state, **which, as we said, is just the standard definition of the overall average value**.

Now, in this case, if one constructs a density operator that is given by the expression

$$\hat{\rho} = P_1 |\phi_1\rangle \langle\phi_1| + P_2 |\phi_2\rangle \langle\phi_2| + P_3 |\phi_3\rangle \langle\phi_3| + \cdots + P_n |\phi_n\rangle \langle\phi_n|$$

sum of probability x projection operators for all possible states

then the expectation value, in this case, is still given by

$$\langle \hat{O} \rangle = \mathrm{Tr}(\hat{\rho}\hat{O})$$

Proof(for mathematically inclined):

Now
$$\langle\phi_1| \hat{\rho} |\phi_1\rangle = \langle\phi_1| (P_1 |\phi_1\rangle \langle\phi_1| + P_2 |\phi_2\rangle \langle\phi_2| + \cdots) |\phi_1\rangle$$

or

$$\langle\phi_1| \hat{\rho} |\phi_1\rangle = P_1 \langle\phi_1 | \phi_1\rangle \langle\phi_1 | \phi_1\rangle + P_2 \langle\phi_1 | \phi_2\rangle \langle\phi_2 | \phi_1\rangle + \cdots$$

but
$$\langle\phi_i | \phi_j\rangle = \delta_{ij}$$

so
$$\langle\phi_1| \hat{\rho} |\phi_1\rangle = P_1$$

similarly

$$\langle\phi_i|\,\hat{\rho}\,|\phi_j\rangle = P_i\delta_{ij}$$

then

$$\mathrm{Tr}(\hat{\rho}\hat{O}) = \sum_i \langle\phi_i|\,(\hat{\rho}\hat{O})\,|\phi_i\rangle = \sum_i \langle\phi_i|\,(\hat{\rho}\hat{I}\hat{O})\,|\phi_i\rangle = \sum_i \langle\phi_i|\,(\hat{\rho}\sum_j(|\phi_j\rangle\langle\phi_j|)\hat{O})\,|\phi_i\rangle$$

or

$$\mathrm{Tr}(\hat{\rho}\hat{O}) = \sum_j\sum_i \langle\phi_i|\,\hat{\rho}\,|\phi_j\rangle\langle\phi_j|\,\hat{O}\,|\phi_i\rangle$$

Thus

$$\mathrm{Tr}(\hat{\rho}\hat{O}) = \sum_j\sum_i P_i\delta_{ij}\langle\phi_j|\,\hat{O}\,|\phi_i\rangle$$

or

$$\mathrm{Tr}(\hat{\rho}\hat{O}) = \sum_i P_i\langle\phi_i|\,\hat{O}\,|\phi_i\rangle = \langle\hat{O}\rangle$$

proof complete!

Some notation: when the density operator takes the form $\hat{\rho} = |\phi\rangle\langle\phi|$ only one term!

it is said to represent a **pure state** and when the density operator takes the form

$$P_1|\phi_1\rangle\langle\phi_1| + P_2|\phi_2\rangle\langle\phi_2| + P_3|\phi_3\rangle\langle\phi_3| + \cdots + P_n|\phi_n\rangle$$

it is said to represent a **mixed state** (more about mixed states later).

A crucial example of the need for the density operator. Follow this very carefully.

Consider a box containing a very large number of electrons, each having spin = 1/2.

This means that each electron spin can have a measurable component ±1/2 along **any** direction.

Now, suppose the box has a hole so that the electrons can get out and then go into a Stern-Gerlach device oriented to measure the z-components of spin(**arbitrary choice of direction**).

In order to proceed, we need to know how the box of electrons was **prepared**, i.e., what state the electrons are in or what the KET labels are…….

Let us consider **two very different cases**:

In the **first** case, we fill the box with electrons that have been prepared in a superposition state

$$|\psi\rangle_1 = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle + |\downarrow_z\rangle) = |\uparrow_x\rangle$$

This preparation can be done by sending electrons through an x-oriented magnet and choosing one of the resulting beams (UP in x in this case).

Thus, in this case, **each** electron is in the the indicated superposition state of z-directions —-> **each** electron is in a superposition of “up” and “down” in the z-direction while **simultaneously** being in a definite state of spin in the x-direction (up). x-value is known and z-value is 50-50

We then fill one box with these electrons.

In the second case, we send electrons through a z-oriented magnet and collect electrons from both beams “z-up” and “z-down” separately.

In this case we **know** that the electrons are either “z-up” OR “z-down”, i.e., they each **have** a definite value - they are **not** in a superposition.

We then fill another box with the two collections of electrons of definite z-spin electrons
Box is now (50-50) up/down in z-direction, **BUT each electron now has a definite value!**.

Remember, the electrons in the box in this case are EITHER in the state $|\uparrow_z\rangle$ OR in the state $|\downarrow_z\rangle$

Now we proceed with the experiments.

In case (1) we observe “z-up” 50% of the time and “z-down” 50% of the time

and we know that in order to describe this system by a state vector we must say

$$|\psi\rangle_{BOX(1)} = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle + |\downarrow_z\rangle) \rightarrow \text{50-50 up-down}$$

i.e., every electron in the box is in this superposition state. which is how we created them

In case (2)) we also observe “z-up” 50% of the time and “z-down” 50% of the time.

But we now have a problem because we are using state vectors as the fundamental object in the theory!

If I did not already know that each electron in the box in this case had a definite value, I would be tempted to describe this system by the same state vector as in case (1).

However, we know that is not true!

The electrons in case (2) are not each in a superposition — they all have **definite values**!

So, if I only measure z-components I cannot tell whether I have case (1) or case (2) and I **do not know how to write the state vector** for the box in case (2),

But remember how I created the electrons in case (1).

They all have a **definite** value of the x-component, namely, “x-up’.

So if I subject the electrons coming out of the box in case (1)

to an x-measurement instead of a z-measurement,

I will end up with only **one beam**!

However, in case (2), the electrons coming out of the box are either “z-up” OR “z-down”

each of which is 50-50 in the x-direction and thus I would end up with two beams after the extra x-measurement!

I get different results for 2nd measurement because they are different states!!

The different results mean that their states must be described differently in QM.

State vectors do not give us the freedom to do this unless we want to monkey around with relative phases between components, i.e., mathematically, we would need to write

$$|\psi\rangle_{BOX(1)} = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle + |\downarrow_z\rangle) \quad , \quad |\psi\rangle_{BOX(2)} = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle + e^{i\alpha}|\downarrow_z\rangle)$$

where α is a **completely unknown relative phase factor**, which must be averaged over during any calculations since it is different (and random) for each separate measurement (each member of ensemble).

Thus, we would be choosing to allow the two vectors to differ in the relative phase of their components and we are forced to say this is meaningful!

Let us continue…

If we use density matrices, we have a very **different story**.

For the pure state a density operator (or matrix) is defined as

$$\hat{\rho} = |\psi\rangle\langle\psi|$$

for some state vector |ψ⟩, i.e., it is the pure state projection operator.

In case (1) this gives $\hat{\rho} = |\psi\rangle\langle\psi| = \frac{1}{2}(|1/2\rangle + |-1/2\rangle)(\langle 1/2| + \langle -1/2|)$

or $\hat{\rho} = \frac{1}{2}(|1/2\rangle\langle 1/2| + |-1/2\rangle\langle 1/2| + |1/2\rangle\langle -1/2| + |-1/2\rangle\langle -1/2|)$ **—> 4 terms**

Derivation of the $\hat{\rho}$ matrix in the (+1/2, –1/2) basis: from our earlier definitions we have

$$\rho = \begin{pmatrix} \langle 1/2|\,\hat{\rho}\,|1/2\rangle & \langle 1/2|\,\hat{\rho}\,|-1/2\rangle \\ \langle -1/2|\,\hat{\rho}\,|1/2\rangle & \langle -1/2|\,\hat{\rho}\,|-1/2\rangle \end{pmatrix}$$

Now

$$\langle 1/2|\,\hat{\rho}\,|1/2\rangle$$

$$= \langle 1/2|\,\frac{1}{2}(|1/2\rangle\langle 1/2| + |-1/2\rangle\langle 1/2| + |1/2\rangle\langle -1/2| + |-1/2\rangle\langle -1/2|)\,|1/2\rangle$$

$$= \frac{1}{2}$$

and so on and we get

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

where the diagonal matrix elements represent probabilities(50-50) and the non-zero off-diagonal matrix elements indicate that one **will observe** quantum interference effects in this system.

It is clear(look at above expression) that any pure state density operator **cannot** be written as the sum of pure state projection operators (multiplying always generates cross-terms).

In case (2), however, each electron is in a definite state (has a definite value) so that each state is separately represented by a single projection operator —>

only have the sum of two definite terms

$$\hat{\rho} = \frac{1}{2}(|1/2\rangle\langle 1/2| + |-1/2\rangle\langle -1/2|)$$

we call this form the "ignorance" density operator

and

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

same algebra

which clearly **is** the sum of pure state projection operators.

This corresponds to what is called a **mixed state**.

Note that the off-diagonals are zero so that this density operator **cannot** lead to any quantum interference effects. The electron **have** values like classical particles!!

Remember for later, this system(case (2)) is such that electrons have values so that the density operator takes this form (sum of projection operators for each value —> NO cross-terms)!!

Note that when electrons DO NOT HAVE VALUES (case (1)) the density operator has interference terms(mixed terms) and cannot be written as a sum of projection operators

A DIFFERENCE that does NOT show up when using state vectors!!

We note that if we treat case (2) as pure state with the extra relative phase factor we would obtain:

$$\hat{\rho} = \frac{1}{2}\begin{pmatrix} 1 & e^{i\alpha} \\ e^{-i\alpha} & 1 \end{pmatrix}$$ which becomes $$\hat{\rho} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

when we average over α .

For comparison, let us digress to see what happens in a **real classical system**.

Now, we digress to see what happens in a real classical system.

Consider rolling a standard die which has possible values = 1,2,3,4,5,6 where the probability of occurrence of each value = 1/6.

In this case, the density operator representing the die will be for a mixed state (no interference effects) and the die **has** a value before/after each roll so that we have

$$\rho = \frac{1}{6}(|1\rangle\langle 1| + |2\rangle\langle 2| + |3\rangle\langle 3| + |4\rangle\langle 4| + |5\rangle\langle 5| + |6\rangle\langle 6|) \qquad \textbf{a mixed state}$$

and the expectation value of the $\widehat{\text{ROLL}}$ operator must be

$$\langle \widehat{ROLL} \rangle = \frac{1}{6}(1+2+3+4+5+6) = 3.5 \qquad \textbf{standard definition}$$

just sum of values times probabilities!!

More formally, we know that any operator can be written as sum of eigenvalues × projections operators, i.e., for the $\hat{B}$ operator introduced earlier we found we can write

$$\hat{B} = \sum_i b_i |b_i\rangle\langle b_i|$$

so that we have for the $\widehat{\mathrm{ROLL}}$ operator

$$\widehat{ROLL} = 1\,|1\rangle\langle 1| + 2\,|2\rangle\langle 2| + 3\,|3\rangle\langle 3| + 4\,|4\rangle\langle 4| + 5\,|5\rangle\langle 5| + 6\,|6\rangle\langle 6|$$

The operator product of the density operator and the $\widehat{\mathrm{ROLL}}$ operator can be written(using the orthonormality of the basis state vectors) as

$$\hat{\rho}\,\widehat{ROLL} = \frac{1}{6}(1\,|1\rangle\langle 1| + 2\,|2\rangle\langle 2| + 3\,|3\rangle\langle 3| + 4\,|4\rangle\langle 4| + 5\,|5\rangle\langle 5| + 6\,|6\rangle\langle 6|)$$

Thus, the expectation or average value is

$$\langle\widehat{ROLL}\rangle = \mathrm{Tr}(\hat{\rho}\widehat{ROLL}) = \sum_k \langle k|\,\hat{\rho}\widehat{ROLL}\,|k\rangle = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

as we obtained earlier. I can do all detailed algebra if interested after class session.

But, in this case, **we know the values are real before the measurement**,

i.e., we are using a macroscopic die with numbers we can see!

Thus, it **seems** that this **particular** form of a density operator **represents that case**.

The same was true earlier for the electrons in the box when we knew they were either “up” or “down”! **We also had a density operator of the same form!**

In both cases the system **had known values before measurements**!

REMEMBER this fact for later.

If we were to add the “collapse” rule it raises a host of questions:

What exactly do we mean, physically and mathematically,

by a “collapse during measurement” of quantum system?

Does collapse occur all at one instant? What if the state occupies a finite volume?

Wouldn’t instantaneous collapse(the entire volume changes at once) contradict special relativity?

If, instead, collapse occurs during a time interval, then what equation describes its time-evolution during that interval?

Quantum states are presumed to follow the Schrödinger equation,

which prescribes continuous time evolution;

How can instantaneous state collapse be reconciled with smooth evolution?

How can we resolve “problem of outcomes” that appears to arise

when a superposed quantum’s state is measured by “which-state” detector,

creating a so-called entangled state of the quantum and the detector

that appears to be an indefinite superposition

of two macroscopically distinct states of a composite system?

Such questions and more comprise the **quantum measurement problem**.

There are many alternative interpretations of the quantum physics mathematical formalism,
and several alternative modifications of theory,
have been proposed to resolve problem, with no consensus on a solution.

It is remarkable that, despite the unparalleled
experimental success of quantum theory across a vast range of experiments,
most of the suggested solutions **differ** from standard quantum physics
in one or more significant aspects.

Most involve **new** interpretations of the standard mathematical formalism.

Interpretations such as "human minds collapse the quantum state"
or "all possible collapses occur but only one of them occurs in our particular universe"
or even rejection of physical reality of quantum world
and assumption that quantum probabilities
(and hence changes in those probabilities,
such as quantum state collapse) are mere measures of personal degrees of belief.

yadda, yadda, yadda……

Other suggestions assume modifications of the standard mathematical formalism,

such as an additional mechanism

that causes quantum states to spontaneously collapse from time to time,

or new "hidden" and hence uncontrollable variables

that create the illusion of quantum randomness. **more yadda, yadda …..**

With all these thoughts/ideas in hand,

we will now continue this discussion of the measurement problem

and finally suggest a resolution of problem of definite outcomes

that lies **entirely** within standard quantum physics.

We will follow a tortuous path.

First, we will spend time with one possible solution to the measurement problem.

We will arrive at a point where it “looks” like we found a solution.

But it will not be so!

The reasons will be subtle but clear.

We will then see how to correctly interpret what is happening and find in the end that von Neumann was correct in 1932 even though for 90 years papers were published deriding his idea!

Begin with question of definitions.

What do we mean, physically, by **quantum measurement**?

First, we need a non-anthropomorphic definition of the concept known as "measurement,"

ie., there were measurements before humans,

so let’s broaden its physical definition as follows:

A “quantum measurement” means

any quantum process that results in a macroscopic effect,

regardless of whether humans or laboratories are involved.

Thus not only is an electron striking a laboratory viewing screen

and creating a visible flash a measurement, but also a cosmic-ray muon striking

and macroscopically moving a sand grain on a planet in some other galaxy

is a measurement.

To analyze a measurement,

we look at a specific experiment: suppose an electron beam passes

through a pair of double slits and then impacts a viewing screen.

Just as in Thomas Young's similar double-slit experiment using light, performed in 1801,

a pattern is formed on the viewing screen that **seems** to shows interference

between the two portions of the electron beam which are **seemingly**(classically)

coming simultaneously through the two slits:

a broad dark-and-bright striped pattern spreads out widely on the screen

- much wider than slits

- indicating regions of destructive (dark) and constructive (bright) interference.

or maybe I should say …

regions of increased electron intensity and decreased electron intensity(including 0)

On closer inspection,

the bright lines are formed by a very large number of tiny individual electron impacts,

each one making a small flash on the screen.

According to our definition, each flash is a measurement

of the position of an electron as hits the screen.

I repeat - Each electron's flash on screen is a measurement!

For the purposes of this analysis, however,

it is more informative to consider a related example of measurement,

still based on the double-slit experiment.

Suppose an electron detector is installed at the slits

and assume that the detector can detect the electron's position

as it passes through slits while disturbing each electron only minimally .

As it turns out, measurement,

even by such a minimally-disturbing "which-path detector", changes everything.

Exactly when the detector turns on,

the pattern on screen changes from a striped interference pattern

to a smoothly-spread-out sum of two single-slit patterns,

each showing diffraction but no interference, i.e., the probability function governing

the pattern of flashes changes at that instant!

The interference pattern abruptly vanishes, i.e., probability has changed.

An analogous experiment has been done using light (photons) instead of electrons,

and using an interferometer rather than double-slit interference setup as we discussed earlier.

A which-path detector was randomly switched on or off

as each photon passed through this experiment;

Photons for which the detector was "off" formed an interference pattern

while photons for which the detector was "on"

formed the expected no-interference pattern.

Reminder:

If the system is in a “pure” superposition state, then the density operator takes the form

$$\hat{\rho} = \frac{1}{2}(|1/2\rangle\langle 1/2| + |-1/2\rangle\langle 1/2| + |1/2\rangle\langle -1/2| + |-1/2\rangle\langle -1/2|)$$

i.e., sum of projection operators and “cross-terms” (interference terms) or the matrix form

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

where the diagonal matrix elements represent probabilities

and the off-diagonal matrix elements imply

that one will observe quantum interference effects in this system.

Clearly, any pure state density operator cannot be written

as the sum of pure state projection operators (because of the cross-terms).

If in a “mixed state”, then the density operator takes the form

$$\hat{\rho} = \frac{1}{2}(|1/2\rangle\langle 1/2| + |-1/2\rangle\langle -1/2|)$$

or the matrix form

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

i.e., sum of projection operators (no interference terms).

As we found in the dice example:

$$\hat{\rho} = \frac{1}{6}(|1\rangle\langle 1| + |2\rangle\langle 2| + |3\rangle\langle 3| + |4\rangle\langle 4| + |5\rangle\langle 5| + |6\rangle\langle 6|) \quad \text{mixed state}$$

We now state our first attempt at describing what happens during a measurement.

When we observe a quantum system evolve into such a “mixed state” density operator,

then the quantum system can be interpreted “classically”,

i.e., a measurement has taken place

and it has been **irreversibly recorded somewhere**.

That is the important point!!

A full “collapse” is not necessary!!

We only need to arrive at (evolve to) a particular form of the density operator!!

Before continuing let me quote Richard Feynman about what has happened.

"Somewhere in the measuring apparatus changes have occurred and the concept of a quantum amplitude is not valid anymore, i.e., if we are throwing our quantum die, then the quantum amplitude only remains a valid idea until the die comes to rest on the floor.

And that point you can't do an experiment which distinguishes interfering alternatives from just plain odds (like with classical dice).

—> we are at the "ignorance" density operator stage!

There is no point where you can say a "reduction" has taken place, i.e., there is just a murky unseen phasing out of the quantum amplitudes.

Possibly what is happening is that in the macroscopic measuring device, quantum phase information gets so "smeared out" that it is no longer definite and thus the system behaves classically.

It registered an irreversible result!

Since we are not quantum objects we may never be able to see the details of this process, but it clearly happens."

Now let us continue our discussion!

Possible ways to think about getting to the irreversible recording point must be found.

Two digressions —> possible ways to get to the irreversible recording —

(1) The so-called Gambler's Ruin problem

We start by investigating the **Mathematical Problem of the Points** (seems unrelated but it is not!)

A sequence of fair coins is flipped.

Player A gets a point for every head and player B gets a point for every tail.

Player A wins if there are a heads before b tails, otherwise B wins.

Find the probability that A wins

Let $\alpha(a,b)$ be the probability that A wins and

$\beta(a,b)$ the probability that B wins —> $\alpha(a,b) + \beta(a,b) = 1$ someone wins!

Solution (due to Pascal and Huygens -1500's)

$$\alpha(a,b) = \frac{1}{2^{a+b-1}} \sum_{k=0}^{b-1} \binom{a+b-1}{k} \quad \text{where} \quad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

binomial coefficient

An example is the **Gambler's Ruin problem**.

$$k! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot k$$

Just as in the problem of points, suppose that at some stage A has a counters,

and B has m + n – a counters so that total number of counters = m+n,

and let A's chances of victory at that point be v(a).

The solution when A starts with m counters is given by

$$p_A = v(m) = \frac{1-(\beta/\alpha)^m}{1-(\beta/\alpha)^{m+n}}$$

This works for all cases except $\alpha = \beta = 1/2$ (the 50-50 probability case).

For that special case, the solution is

$$v(a) = \frac{a}{m+n}$$

This implies that there is always a winner (hence the the name **Gambler's Ruin** for the loser)

NOTE: Even when $\alpha = \beta = 1/2$ there is still always a winner!!!!

Also note that this relation is **linear** in **a** in this special case!

Pearle's Theorem

What does this have to do with quantum collapse?

Here is a short note from Philip Pearle (a really good physicist and friend): I quote:

I soon found a charming **analogy** for collapse dynamics,

useful for providing an intuitive and non-technical **explanation of how it works**.

I happened to be browsing in Feller's book on probability

(a favorite textbook, from an undergraduate course taught by Stanislaus Ulam)

when I encountered the **gambler's ruin game**.

Two gamblers, initially possessing, respectively, a **fraction** $x_1(0)$, $x_2(0)$ of their combined wealth

(so $x_1(0) + x_2(0) = 1$)

repeatedly toss a fair coin, and the result, heads or tails,

determines which one gives one dollar to the other.

They play until one gambler loses all his money, and the game ends.

The **analogy** arises if the amount of money possessed by one gambler at any time

is proportional to the squared amplitude of one of two states

whose sum is the state vector representing the physical system undergoing collapse.

Just as one gambler loses all his money, **so one of the states loses all its amplitude**,

and as the other gambler wins all the money,

so the state vector ends up as totally described by the other state (—> collapse!!).

collapse has occurred without any interference by gamblers!

Let Q(x) be the conditional probability that a gambler wins the game,

given that he has the fraction x of the total wealth.

If Δ is the fraction of the total wealth they exchange at each toss (i.e., Δ = \$1/total dollars),

the difference equation

$$Q(x) = \frac{1}{2}Q(x - \Delta) + \frac{1}{2}Q(x + \Delta)$$

expresses that there are two routes to win if one has fractional wealth x,

namely lose the next toss and drop to x − Δ but win thereafter,

or win the next toss and rise to x + Δ and win thereafter.

The solution of the difference equation is Q(x) = Ax + B, where A and B are constants.

Boundary conditions:

Since Q(0) = 0 (because you can't win if you have no money) and Q(1) = 1

(because you have won if you have all the money), then Q(x) = x, i.e.,

Q(1) = 1 = A+B Q(0) = B= 0 —> A = 1, B = 0 —> Q(x) = x

That is, if one starts with the fraction x = x(0) of the money,

one has the probability x(0) of attaining all the money —> you get to x = 1,

which is exactly collapse behavior.

The game can be modified to have many players,

to have Δ change as the game progresses

(e.g., to get smaller as one gambler gets closer to losing), etc.

So, one may think of quantum collapse

(getting to the point where an irreversible measurement has been made)

as a gambler's ruin competition among the states in a superposition,

to see which final state wins the game (remains at the end).

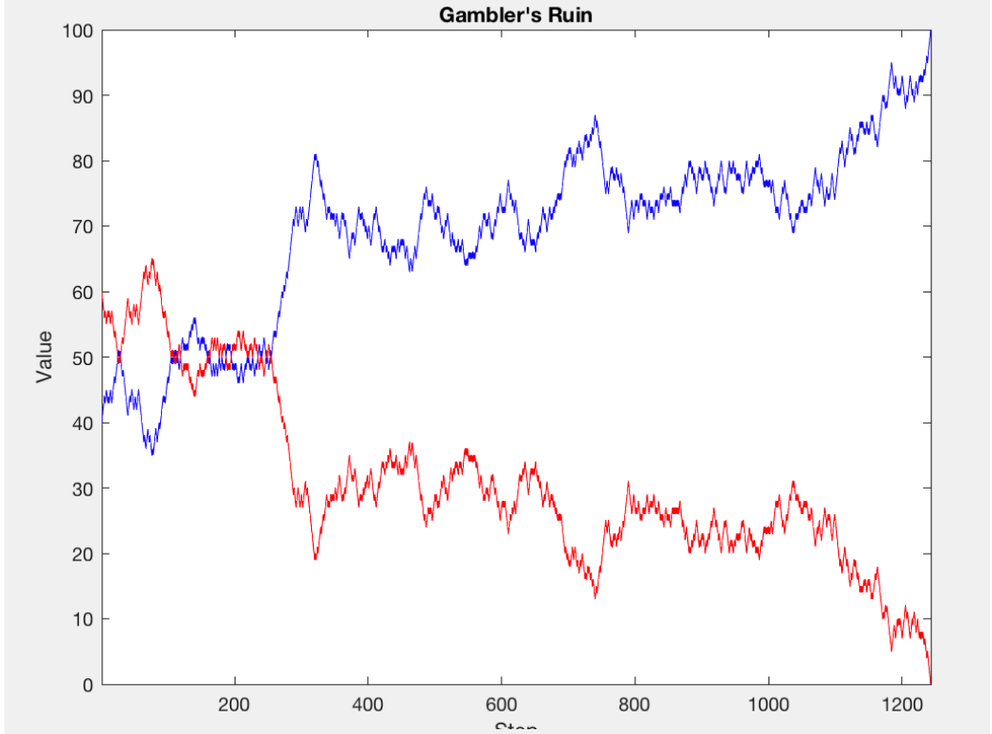
I ran a simulation of the Gambler's Ruin problem using OCTAVE on a computer.

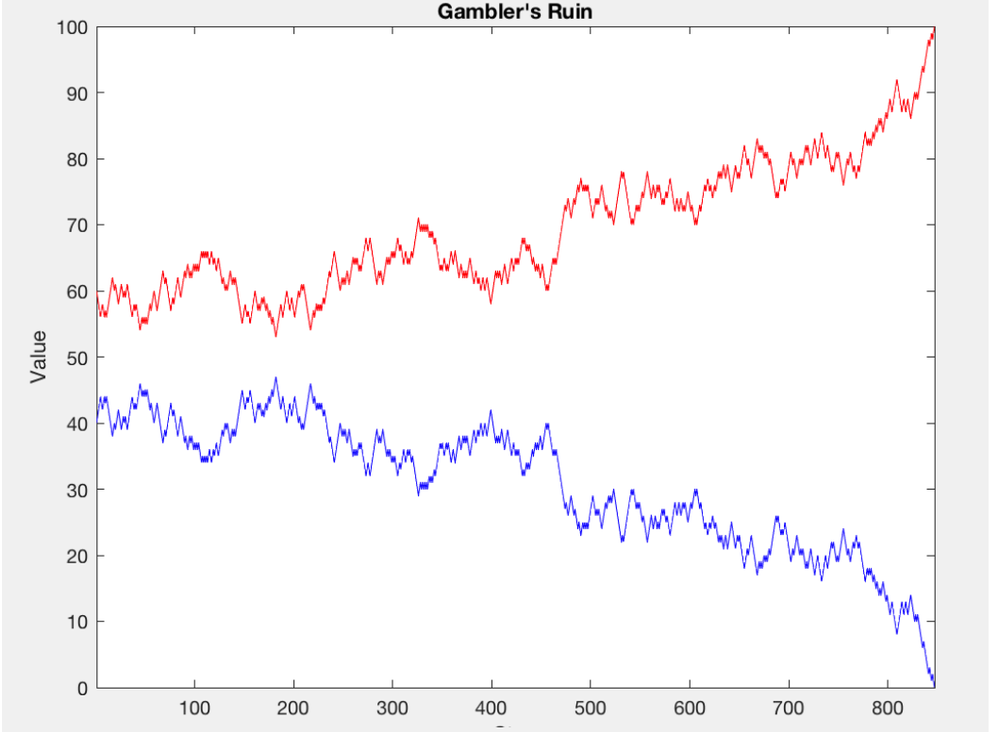
OCTAVE Code

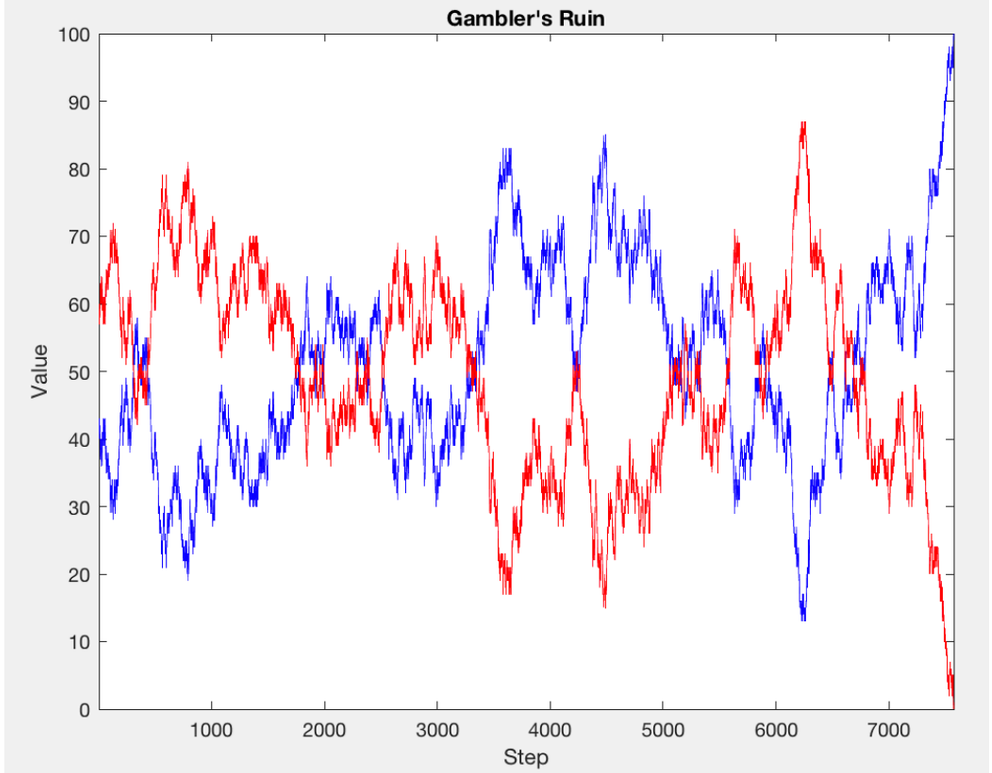
gambruin.m

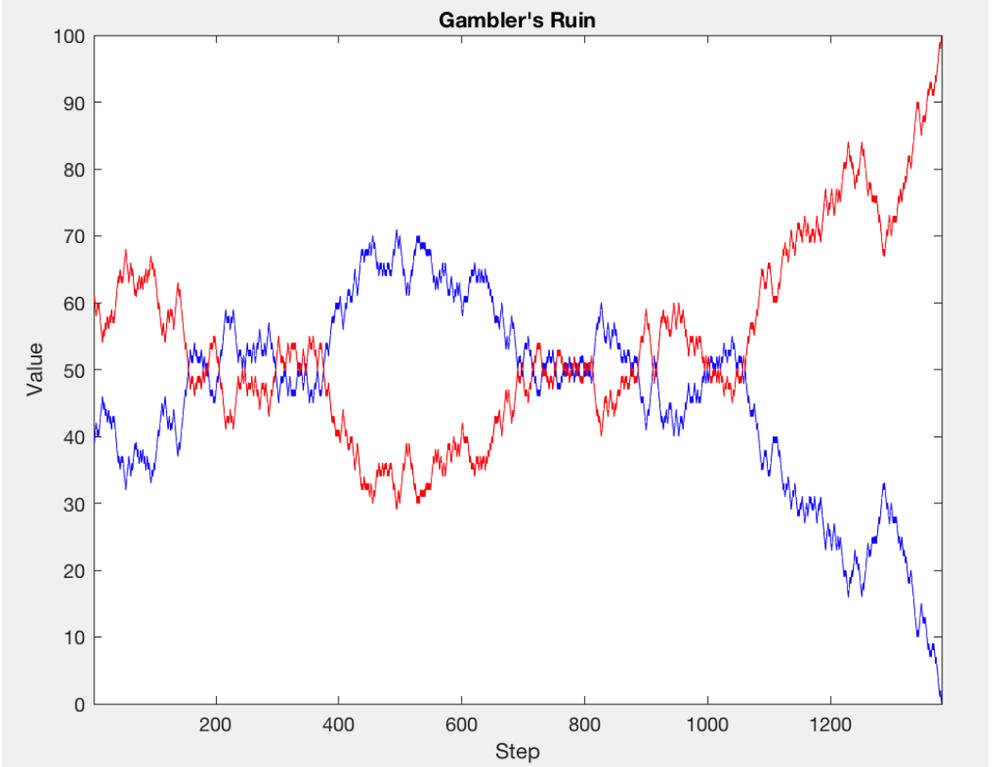
```
x1=[];
y1=[];
x=40;x1=[x1,x];
y=100-x;y1=[y1,y];
count=1;
while (x ~= 0) && (x ~= 100)
r=rand;
x=x+(r <= 0.5)-(r > 0.5);x1=[x1,x];
y=100-x;y1=[y1,y];
count=count+1;
endwhile
figure
plot(1:count,x1,'-b');
axis([1,count,0,100]);
xlabel('Step');
ylabel('Value');
title('Gambler''s Ruin')
hold on;
plot(1:count,y1,'-r');
hold off
count
```

Some sample runs are shown in next slide - **note there is always a winner(loser)**









One state always wins out!!

Collapse occurs even though there is no direct collapse mechanism!!

An irreversible mark always appears!!

Could this be the way collapse works without a postulate or mechanism?

More Details are in supplementary reading GamblerRuin.pdf

Another way to get to the irreversible recording ……

(2) Decoherence…

If we have a quantum system is the state

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

then this(non-zero off-diagonal elements) states **exhibits** quantum interference effects.

"Decoherence" says that as a state interacts with macroworld in its environment (which it has to do) the off diagonal elements go to zero, i.e., the state makes transition to

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

As we already said, this state has zero quantum interference effects and a definite measurement value has been irreversibly recorded somewhere.

If we had quantum dice the state vector would be

$$|Dice\rangle = \frac{1}{\sqrt{6}}(|1\rangle + |2\rangle + |3\rangle + |4\rangle + |5\rangle + |6\rangle) \quad \text{and} \quad \rho = |Dice\rangle\langle Dice| \quad \textbf{giving}$$

$$\begin{aligned}\rho = \frac{1}{6}[&|1\rangle\langle 1| + |2\rangle\langle 1| + |3\rangle\langle 1| + |4\rangle\langle 1| + |5\rangle\langle 1| + |6\rangle\langle 1| \\ &+ |1\rangle\langle 2| + |2\rangle\langle 2| + |3\rangle\langle 2| + |4\rangle\langle 2| + |5\rangle\langle 2| + |6\rangle\langle 2| \\ &+ |1\rangle\langle 3| + |2\rangle\langle 3| + |3\rangle\langle 3| + |4\rangle\langle 3| + |5\rangle\langle 3| + |6\rangle\langle 3| \\ &+ |1\rangle\langle 4| + |2\rangle\langle 4| + |3\rangle\langle 4| + |4\rangle\langle 4| + |5\rangle\langle 4| + |6\rangle\langle 4| \\ &+ |1\rangle\langle 5| + |2\rangle\langle 5| + |3\rangle\langle 5| + |4\rangle\langle 5| + |5\rangle\langle 5| + |6\rangle\langle 5| \\ &+ |1\rangle\langle 6| + |2\rangle\langle 6| + |3\rangle\langle 6| + |4\rangle\langle 6| + |5\rangle\langle 6| + |6\rangle\langle 6|]\end{aligned}$$

Thus, the quantum dice generally has a lot of quantum interference terms.

This corresponds to a density matrix of the form

$$\rho = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{pmatrix}$$

lots of off-diagonal terms

lots of quantum interference

Rolling the dice (a measurement) or “decoherence” if dice left alone, produces

$$\rho = \begin{pmatrix} a_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{66} \end{pmatrix}$$

where $a_{11} = a_{22} = a_{33} = a_{44} = a_{55} = a_{66} = \frac{1}{6}$

or

$$\hat{\rho} = \frac{1}{6}(|1\rangle\langle 1| + |2\rangle\langle 2| + |3\rangle\langle 3| + |4\rangle\langle 4| + |5\rangle\langle 5| + |6\rangle\langle 6|) \quad \text{mixed state}$$

and as we said earlier this indicates a measurement has occurred!! .

Maybe that is what Gambler’s Ruin is doing and they are the same thing!

Now let us continue on our earlier path.

Now look closely at “which-path” experiments

In so-called "delayed-choice experiment” mentioned earlier,

there is instantaneous (to within some accuracy) fast switching between two states;

Also, each collapse is executed entirely while photon was inside interferometer.

We can gain considerable insight by studying

how quantum theory describes a which-path measurement.

Note: it is a measurement as defined earlier,

because a detector **registers** "slit 1" or "slit 2" macroscopically for each electron.

Denote state of one electron passing through slit 1 as $|\psi_1\rangle$

and the state of one electron passing through slit 2 as $|\psi_2\rangle$.

John von Neumann, who was the first to

carefully analyze measurement in purely quantum-theoretical terms,

insisted on treating not only the measured quantum but also the macroscopic detector as quantum systems because, after all, detectors are made of atoms and they perform a quantum function by detecting individual quanta.

I agree with that view of the macroscopic detector.

Now repeat earlier discussion again, filling in any remaining gaps and eliminating any remaining confusions.

Accordingly, represent "ready to detect" quantum state of detector by $|ready\rangle$,

and state of detector after detecting an electron by $|1\rangle$ if $|\psi_1\rangle$ was detected,

and by $|2\rangle$ if $|\psi_2\rangle$ was detected.

A properly operating detector will surely transition from $|ready\rangle$ to $|1\rangle$

upon measurement of electron that has been prepared

(perhaps by simply shutting slit 2) in $|\psi_1\rangle$ state .

As a limiting idealization,

we assume, with von Neumann,

that the measurement of electron prepared in the state $|\psi_1\rangle$

leaves the electron still in state $|\psi_1\rangle$ after detection.

Such a **minimally-disturbing measurement** would cause

the electron-plus-detector composite system,

initially in the composite state $|\psi_1\rangle\,|ready\rangle$,

to transition into the final state $|\psi_1\rangle\,|1\rangle$.

We summarize the process as

$$|\psi_1\rangle\,|ready\rangle \rightarrow |\psi_1\rangle\,|1\rangle \qquad \textbf{(1)}$$

Similarly, a minimally-disturbing measurement of an electron initially prepared in $|\psi_2\rangle$ is described mathematically by

$$|\psi_2\rangle\,|ready\rangle \rightarrow |\psi_2\rangle\,|2\rangle \qquad \textbf{(2)}$$

Now suppose that both slits are open so each electron "can pass through either slit",

and suppose the preparation and experiment (e.g. slit widths) is symmetric

with respect to the two slits.

Then the state of each electron as it approaches the slits prior to detection

must be described by a symmetric superposition

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\psi_1\rangle + |\psi_2\rangle) \qquad \textbf{(3)}$$

But quantum physics, including its time dependence, is **linear**. This implies that $|\psi\rangle\,|ready\rangle$ evolves according to

$$\begin{aligned} |\psi\rangle\,|ready\rangle &= \frac{1}{\sqrt{2}}(|\psi_1\rangle + |\psi_2\rangle)\,|ready\rangle \\ &= \frac{1}{\sqrt{2}}(|\psi_1\rangle\,|ready\rangle + |\psi_2\rangle\,|ready\rangle) \\ &\rightarrow \frac{1}{\sqrt{2}}(|\psi_1\rangle\,|1\rangle + |\psi_2\rangle\,|2\rangle) \end{aligned} \qquad \textbf{(4)}$$

The final state

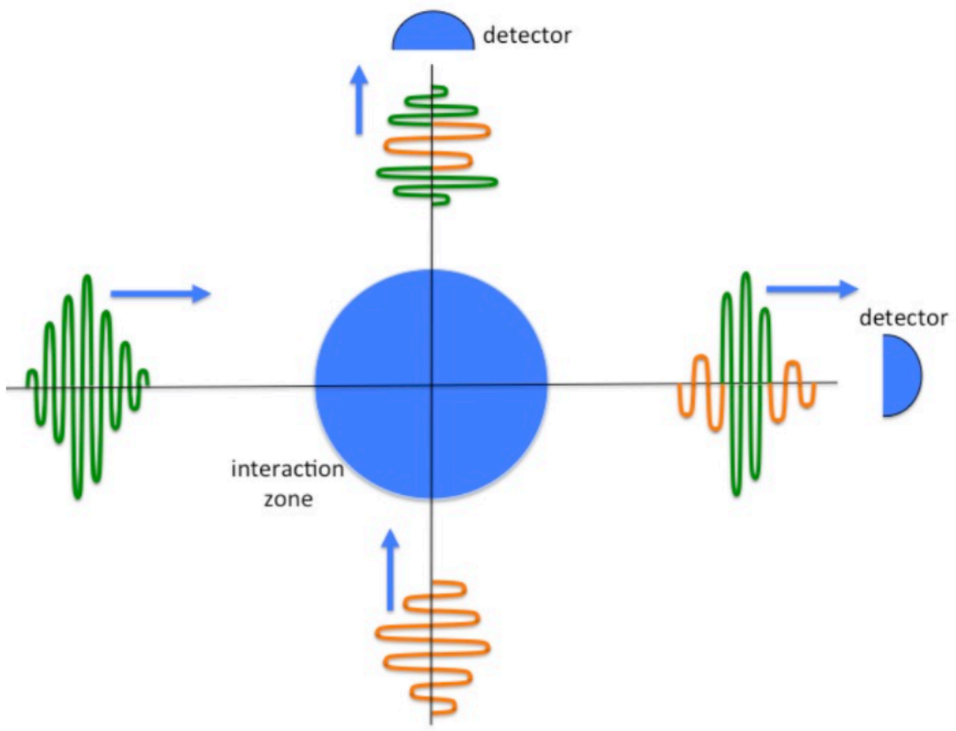
$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\psi_1\rangle\,|1\rangle + |\psi_2\rangle\,|2\rangle) \qquad \textbf{(5)}$$

following detection is said to be **"entangled"**

because it **cannot be factored** into simple product of states of the two sub-systems.

You should make a copy of all the numbered equations for reference as I lecture

As indicated suggestively in Figure



when two independent quanta pass **near** each other, interact, and subsequently separate, the interaction generally **entangles** the two quanta and the entanglement then **persists** after interaction **regardless** of how far apart the two quanta might eventually travel, provided only that the two quanta experience **no further interactions**.

Despite a possibly wide spatial separation, entangled quanta have a **unity** not possessed by non-entangled quanta.

This unity is source of quantum non-locality!

Entanglement is ubiquitous in nature.

The entangled "measurement state" (5) that is at **heart** of quantum measurement is remarkably **subtle**.

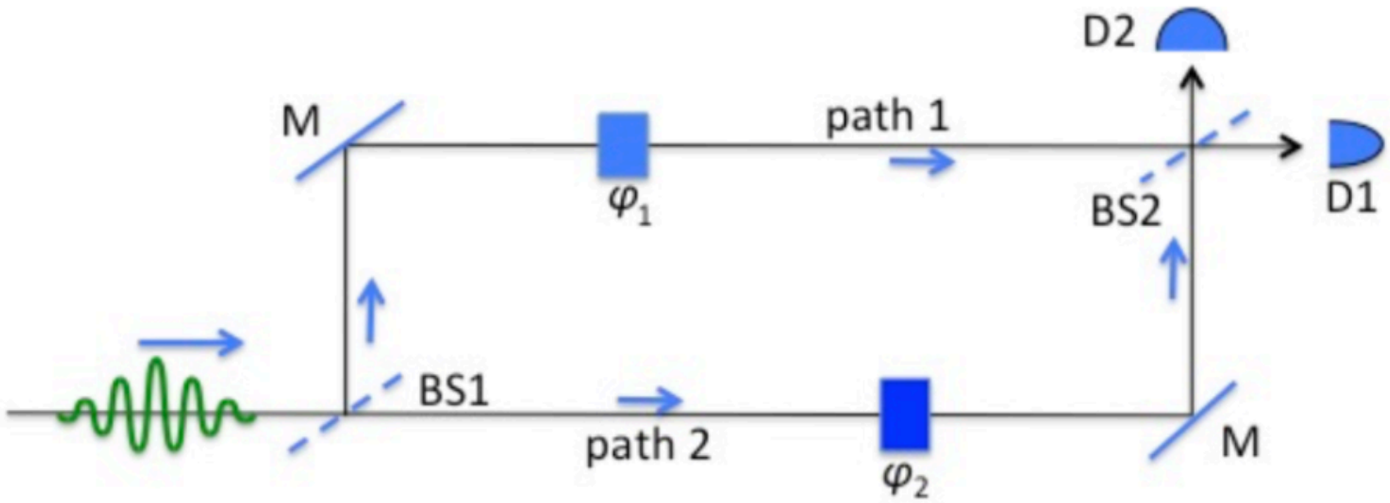
To fully understand **“entanglement”**, we **first** need to understand **“superposition”**.

The quantum principle says that

any linear combination(superposition) of possible quantum states of a system,

as in (3) and (5) for example, is also a possible quantum state of that system.

Figure pictures an experiment that demonstrates such a superposition of states.



This represents a layout of optical paths called a “Mach-Zehnder interferometer”.

Light beam enters at lower left passing through "beam splitter" BS1;

reflected beam(amplitude) makes a right angle with incoming direction

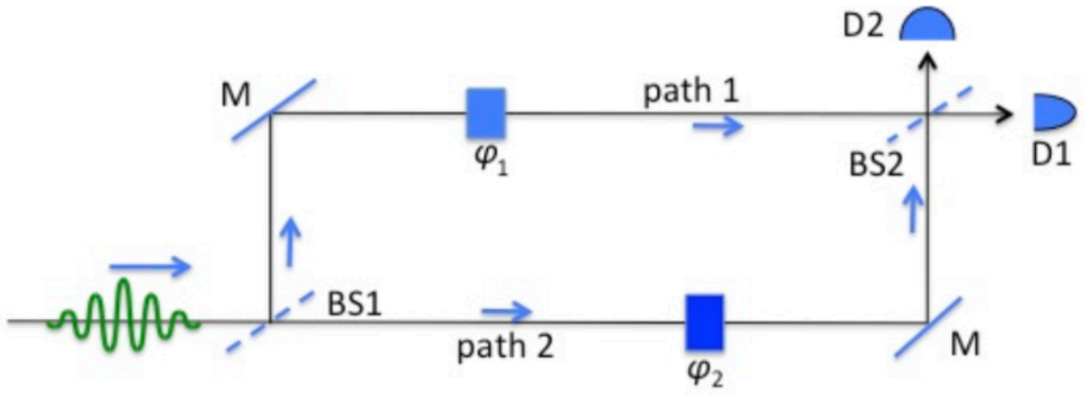
while transmitted beam(amplitude) passes straight through.

So the beam “splits” and each “half” traverses one of two paths (meaning amplitudes exist for such to happen);

Mirrors M bring paths back to crossing point as shown.

Devices called "phase shifters,

" denoted by ϕ_1 and ϕ_2, are placed into each path.



The phase shifter can add

a short variable length to a path.

A second beam splitter BS2 can be placed at the final crossing point.

Without BS2, each “half”-beam(amplitude) moves straight ahead along one path

to the detector on that path.

Things get more interesting with BS2 in place.

Because 50% of each of the two beams(amplitudes) then goes to each detector,

BS2 mixes the two beams(amplitudes) together so the experiment can show interference.

The interferometer is constructed so that, when phase shifters are set to zero,

two "optical paths" (number of wavelengths, after accounting for phase changes upon reflection and refraction) from the entry point to D1 are equal

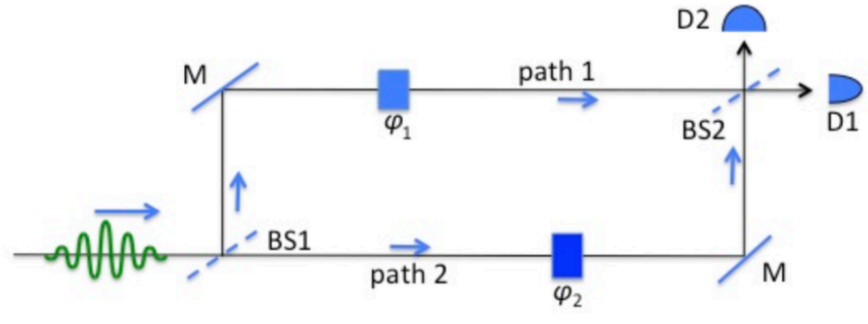
while the two optical paths to D2 differ by half a wavelength.

Thus, is found that amplitudes interfere constructively at D1 and destructively at D2,

i.e., all photons go to D1.

If one then uses ϕ_1 or ϕ_2 to add half wavelength to either path, amplitudes then interfere constructively at D2 and destructively at D1,

i.e., all photons go to D2.



As one continuously varies length of one or other path

by varying one or the other phase shifter,

one finds that amount of light(number of photons) arriving at D1 varies continuously from

100% down to 0%,

while amount arriving at D2 varies from 0% to 100%.

Two paths are clearly interfering -> can construct a normal interference pattern from the data!

Experiment is an interferometer-based analog of Young's double-slit interference experiment demonstrating the "wave" nature of light - because have set the **experiment context** to do so!.

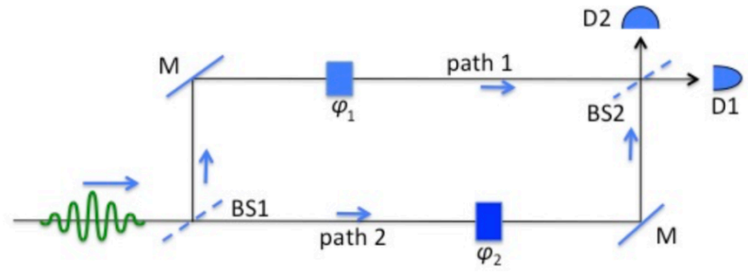
But, as we know, light is really just photons, and photons are indivisible.

How then does **nature** explain this experiment

when we dim the light source to a level where only **one photon at a time** traverses the

interferometer? i.e., when the detectors only record one photon at any instant of time!

After all, the photon still traverses BS1, but it cannot split in two because a **quantum is unified and cannot be split.**



With BS2 removed, one finds either D1 or D2 registers a single entire photon, randomly, with 50-50 probabilities, regardless of how the phase shifters are set.

The randomness is **absolute** - it is more random than any human macroscopic game, such as coin flips, which only mimics randomness.

Nature **has invented** quantum randomness to deal with obstacles such as beam splitters while **preserving the unity of the quantum**.

Detectors **never** register half a photon. They **either** get either a whole photon or no photon.

What happens in single-photon experiment with BS2 present?

As discussed earlier,

beginning from equal path lengths,

which give constructive interference at D1

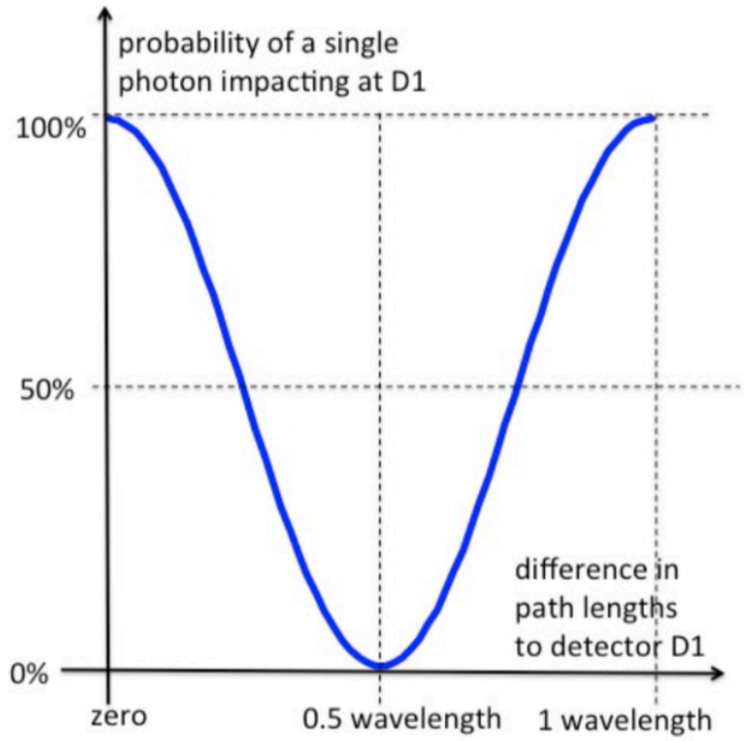
and destructive interference at D2,

as phase shifters vary,

probabilities of detecting a photon at D1 and D2

vary as in Figure,

which gives the percentage of photons impacting D1.



Importantly,

results do not depend on which phase shifter the experimenter chooses to vary.

Since each photon responds to changes in either path length,

each photon **must "follow both paths" -** **whatever that statement means**!

This verifies the superposition principle and shows that quanta can **seem** to be in **"two places at same time".**

This seems **paradoxical** if we **assume** photons are tiny particles,

but if we assume photons are waves it is **not** paradoxical,

i.e, each photon simply spreads along both paths, interfering with itself at D1 and D2.

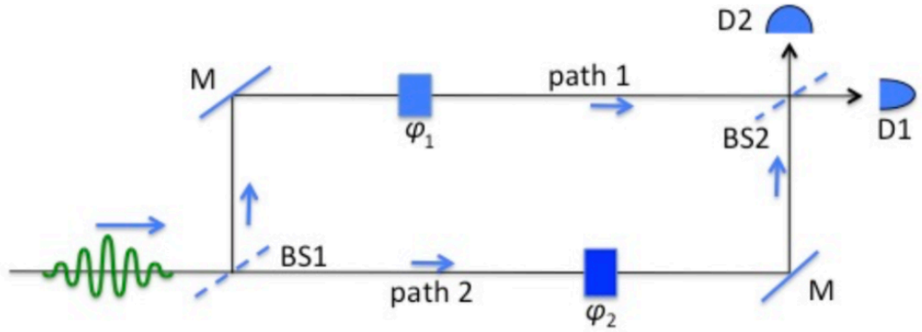
One must conclude that each photon travels both paths

even when BS2 is not present to directly verify this,

because once a photon enters interferometer

it **must behave in same manner** regardless of whether BS2 is placed

or not placed at far end.



A **delayed-choice** experiment provides further evidence for this conclusion:

Since photons **"do not know"** whether BS2 will be inserted,

they must travel both paths on all trials

including those for which BS2 is not inserted - **this is connected with entanglement**.

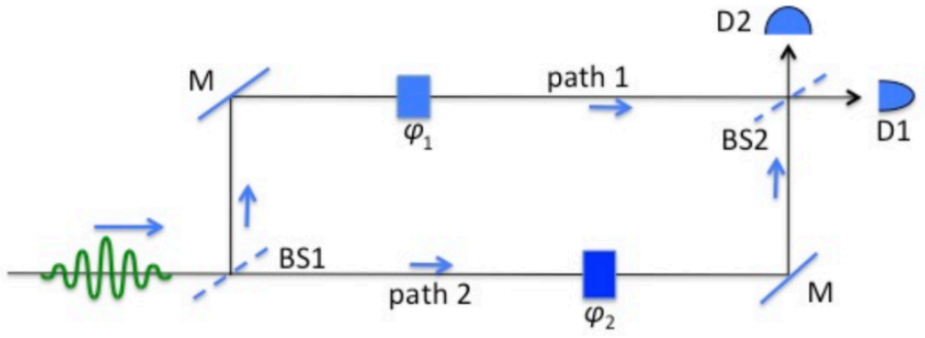
With BS2 removed,

the situation is like the double-slit experiment

with a which-slit detector present.

Each photon is entangled

with macroscopic detectors D1 or D2 as in (5).



With BS2 present, the two paths mix and we have a situation

like double-slit experiment with no which-slit detector -

each photon follows two paths to each detector

where it interferes with itself, and we detect an interference state (3).

All of this suggests that measurements affect superposed quantum states

via entanglement of the superposed quantum with a detector.

Resolving Paradoxes and Understanding Measurement

The apparent paradox of Schrödinger's cat

A cat is penned up in a steel chamber, along with the following setup.

In a Geiger counter there is tiny bit of radioactive substance, so small, that in course of an hour one of its atoms decays, but also, with equal probability, none decay; if it happens, the counter tube discharges and through a relay releases a hammer which shatters a small flask of hydrocyanic acid (poisonous vapor).

If one left the entire system to itself for an hour,

one would say that the cat still lives if no atom has decayed!

If one uses a state vector approach for understanding the entire system,

one would express this by having a living and a dead cat mixed

or smeared out in equal parts -

a superposition of “dead” and “alive” with other properties of system!

a statement that is as stupid as it sounds!!

It is typical of these cases,

that indeterminacy originally restricted to atomic domain

becomes transformed into a macroscopic indeterminacy,

which can only then be resolved by direct observation,

which prevents us from naively

accepting as valid the “blurred model" for representing reality.

Mathematically, the nucleus and the cat

have become entangled in measurement state (5),

with $|\psi_1\rangle$ and $|\psi_2\rangle$ representing the undecayed and decayed states of the nucleus

and $|1\rangle$ and $|2\rangle$ representing the alive and dead cat.

According to Schrodinger's understanding of this situation,

the indeterminacy of the microscopic nuclear state "becomes transformed into the macroscopic indeterminacy" of cat, and he could not comfortably accept this "blurred cat state" i.e., a cat that is in a superposition of being both alive and dead.

He hoped this would say something is wrong with QM.

As we will show, standard quantum physics says that Schrodinger's 1937 understanding was **incorrect**.

The composite system (cat-plus-nucleus) is not predicted to be in superposition of two states of a cat and two states of a nucleus.

Instead, the composite system is predicted to be in a superposition of two **correlations** between the cat and the nucleus; one where live cat is 100% correlated with undecayed nucleus, and second where dead cat is 100% correlated with decayed nucleus.

Entanglement will have transformed a pure state superposition of nuclear states to a pure state superposition of correlations between subsystem states.

We will see this is what one expects from quantum mechanics, and it is **not paradoxical**.

This so-called “problem of definite outcomes”

applies of course to more than Schrodinger’s dramatized example.

Regardless of whether measuring instrument is a which-slit detector,

a Geiger counter, or a cat, the entangled state (5) applies.

This state appears at first glance to represent a quantum superposition

in which the detector is in two macroscopically different states simultaneously.

If so, then there is an inconsistency within quantum physics,

because obviously it **cannot** be this easy to create a macroscopic superposition.

The question is: Is it true that (5) really represents a macroscopic superposition?

Turns out, there is more to this entangled state than meets the eye.

If one assumes the detector to be in a superposed state in its own space $a|1\rangle + b|2\rangle$,

one finds that (5) necessitates either a = 0 or b = 0 (one or the other), implying that the detector is not in an individually superposed state **within its own Hilbert space**.

The **same** applies to the detected quantum:

It is not in a superposed state $a|\psi_1\rangle + b|\psi_2\rangle$ with both $a \neq 0$ and $b \neq 0$.

The entanglement process leaves neither sub-system superposed in **own** space!

So far as I know, this simple fact has long been ignored by physicists studying the measurement problem.

The Density Operator to the Rescue

The density operator formalism for quantum physics provides a stronger version of this conclusion(because it is the more appropriate approach).

Density operator for quantum system whose state is $|\psi\rangle$(pure state) is a projection operator

$$\hat{\rho} = |\psi\rangle\langle\psi| \tag{6}$$

As we saw earlier, if system is in a state whose density operator is $\hat{\rho}$, then standard quantum expectation value $\langle\hat{O}\rangle$ of an arbitrary observable $\hat{O}$ is found from

$$\langle\hat{O}\rangle = \mathrm{Tr}(\hat{\rho}\hat{O}) \tag{7}$$

where "Tr" represents trace operation (sum of diagonal matrix elements).

This approach is especially useful if the quantum system is a composite of two subsystems A and B.

We define the density operator $\hat{\rho}_A$ for subsystem A alone by

$$\hat{\rho}_A = \mathrm{Tr}_B(\hat{\rho}) \tag{8}$$

where "Tr_B" means that trace taken **only** over states of subsystem B (remove all traces of B).

It is then easy to show (see later)

that standard quantum expectation values for subsystem A alone

(values obtained by an observer of A **without** any knowledge of B) are

$$\langle\hat{O}_A\rangle = \mathrm{Tr}(\hat{\rho}_A\hat{O}_A) \tag{9}$$

where $\hat{O}_A$ means any observable operating on system A alone

(i.e., operating within A's Hilbert space).

Applying this formulation to measurement state (5),

the reduced density operators for the quantum system (call it A) and its detector (call it B), respectively, are

$$\hat{\rho}_A = \frac{1}{2}(|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|) \qquad \textbf{(10)}$$

$$\hat{\rho}_B = \frac{1}{2}(|1\rangle\langle 1| + |2\rangle\langle 2|) \qquad \textbf{(11)}$$

Plus signs in (10) and (11) make one think of superpositions such as (3), but these are **not** superpositions.

The density operator for the superposition (3) has cross-terms:

$$\hat{\rho} = (|\psi\rangle\langle\psi|) = \frac{1}{2}(|\psi_1\rangle\langle\psi_1| + |\psi_1\rangle\langle\psi_2| + |\psi_2\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|) \qquad \textbf{(12)}$$

Two cross-terms, involving both $|\psi_1\rangle$ and $|\psi_2\rangle$, are missing in (10).

So (10) does not describe a system in a superposition of two quantum states.

However, (10) is precisely the density operator one should use if one **knows** the quantum system **is either** in state $|\psi_1\rangle$ **or** in state $|\psi_2\rangle$ but one didn't know which and so, due your own to lack of information,

—> you simply assign a probability of 1/2 to each of the two possibilities.

Same goes for (11).

(10) and (11) are **"classical"** probabilistic states - **analogous** to "states of knowledge" one would assign to a coin flip when you know outcome to be either heads or tails with equal probability **but don't know yet** which has occurred.

Remember the dice example also!

The situation described by a density operator such as (10) is known as a "mixture" of states $|\psi_1\rangle$ and $|\psi_2\rangle$, as distinct from a "superposition" of states observed in the Mach-Zehnder experiment and represented by (3).

Equation (9) tells us that all correct statistics for subsystem A alone can be found from standard formula (7) applied to subsystem A **alone**.

But we have just seen that (10) is the density operator one should use if one knows A to be in either $|\psi_1\rangle$ and $|\psi_2\rangle$ without knowing which.

The same goes for subsystem B and (11).

In the case of Schrodinger's cat, it follows that the observer of the cat alone sees outcomes appropriate to a cat that is **either** alive **or** dead, **not** both.

For subsystems, the interference terms are missing, and an "ensemble" of repeated trials **must** exhibit a nonsuperposed mixture rather than a superposition.

This is the clear prediction of quantum physics for the entangled state (5).

But one must be careful, because (10) and (11) are not complete descriptions of the quantum states of the nucleus or the cat.

In fact, (10) and (11) are not quantum states at all, but merely **"reduced states"** arising from the actual state (5) of the composite system when one part of the composite system is removed from the equations..

In the case of Schrodinger's cat, (10) and (11) give the correct predictions for observations of either the nucleus alone or the cat alone, but do not represent the state of either subsystem because this given by (5).

In fact, when two quanta are entangled, neither one has quantum state of its own!

But B's state of affairs is certainly **not entirely** described by (11).

Rather, it is described by the composite state (5).

Equation (11) **merely** tells us following:

If the cat and the nucleus are in state (5) then, when one looks at the cat, one is going to see a cat that is either alive or dead.

There is no claim that (11) represents the complete quantum state of the cat.

That is, there is no claim that the cat is really in either state $|1\rangle$ or state $|2\rangle$, because the state it's really in is admittedly (5).

In fact, we do have complete knowledge of the state of both A and B, namely measurement state (5).

Reduced operators admittedly do not represent the state of the composite system.

They tell us only what will observe at the nucleus and at the cat

and they tell us nothing about the correlations between these observations,

so these density operators do not tell us the real state of the system.

And so plot thickens.

The entangled state (5) properly describes both individual subsystems.

However, the plus sign in (5) signifies the superposition of two terms.

We know, however, that neither subsystem A nor subsystem B is superposed.

What is meaning of plus sign?

This superposition arose from superposition represented by (3).

We cannot logically ignore this fact -

a strategy known as "shut up and calculate" approach to quantum measurement.

Instead, we must ask:

Exactly what is superposed when two subsystems are in this entangled state?

Superpositions preserve the all-important **unity of quantum**.

When Max Planck proposed in 1900 that electromagnetic radiation occurs in energy units of magnitude $E = h\nu$, he tacitly implied the central quantum principle:

The unity of an individual quantum.

Energy (electromagnetic energy in case of radiation) comes in spatially extended bundles, each having a definite and identical quantity of energy.

One cannot have half a quantum, or 2.7 quanta.

You must have either 0 or 1 or 2 etc. quanta.

This is a fairly natural notion - apparently nature prefers to sub-divide the universe into a countable or even a finite set of entities as opposed to an uncountable continuum.

The spatial extension of these bundles **then** implies nonlocality:

If we have one quantum and destroy it (by transforming it to something else),

we **must** destroy all of it **everywhere**,

simultaneously, because we **cannot**, at any time, have **just** part of quantum.

Louis de Broglie put it perfectly in 1924, regarding another kind of quantum, namely the electron:

The energy of an electron is spread over all space with a strong concentration in a very small region. ...That which makes an electron an atom of energy is not its small volume that it occupies in space – I repeat it occupies all space – but the fact that it is indivisible, that it constitutes a unit.

When one transforms state of a quantum, one **must** transform the entire extended quantum all at once.

Hence there are observed quantum jumps in experiments.

Furthermore, composite entangled systems such as atoms **also** behave in a unified fashion.

This unity is the source of nonlocality seen in experiments involving entangled pairs of photons.

Nonlocality is exactly what one would expect, given the unity and spatial extension of quantum and unitary (i.e. unity-preserving) nature of entanglement process.

Standard nonrelativistic quantum theory seems to prescribe two kinds of time evolution:

collapse upon measurement,

and Schrodinger equation between measurements.

The key feature of Schrodinger equation is that it prescribes a so-called "unitary" time evolution,

meaning time evolution that preserves pure states,

i.e. transforms unit Hilbert space vectors into other unit vectors.

Some ideas that are required physically by the unity of quantum are expressed as follows:

If quantum is described by pure quantum state at t = 0, it should **remain** pure at later times.

This notion prompts us to **ask** whether the measurement process also preserves pure states.

At least in case of idealized process described in (4), the answer is **"yes"** because **both** "before" and "after" states are pure.

Measurement state (5), since it is pure, **represents** a highly unified state of affairs, **even** though one of its subsystems is a macroscopic detector.

Thus, one suspects that this state, like its progenitor (3), is truly a superposition in which the superposed terms **represent** two situations or states of the same object.

But precisely what is that object, i.e., **what** is superposed?

We have seen that the states of subsystem A are **not** superposed,

nor are the states of subsystem B.

The conventional interpretation (which, as we will see, is subtly incorrect) of a product state such as $|\psi_1\rangle|1\rangle$ is that it represents a state of a composite system AB in which subsystem A is in state $|\psi_1\rangle$ while B is in state $|1\rangle$ (we used such a state in the two-path experiments).

In this case, (5) would represent a superposition in which AB is simultaneously in the state $|\psi_1\rangle|1\rangle$ and also in the state $|\psi_2\rangle|2\rangle$.

The situation for Schrodinger's cat would be: live cat and undecayed nucleus superposed with dead cat and decayed nucleus.

This is at least as physically outrageous as a live cat superposed with a dead cat, and it contradicts physical implications (a cat that is either alive or dead) of reduced states (10) and (11) as described earlier.

Something is still wrong!

Let us repeat some of our discussion and provide more details to work it all out.

Some Repetition and More Intricate Details

Proposal:

The solution to the so-called quantum measurement problem is completely contained within standard quantum mechanics and needs no elaborate new structures and interpretations.

Remember the standard discussion from earlier:

The following discussion is at limits of knowledge you have learned about QM, but I thought you should see what researchers including myself are doing at the present time. The details of the derivations are not important, ONLY the conclusions.

Total system T = quantum system S (states $|s_i\rangle$) + measuring device A (states $|a_i\rangle$), where $|a_0\rangle = \text{ state of measuring device “off”}$

The unitary time evolution rule then says

$$|s_1\rangle|a_0\rangle \rightarrow |s_1\rangle|a_1\rangle \quad \text{unitary evolution} \qquad |s_2\rangle|a_0\rangle \rightarrow |s_1\rangle|a_2\rangle \quad \text{unitary evolution}$$

The linearity of QM then says that

$$|\psi\rangle_{SA} = (c_1|s_1\rangle + c_2|s_2\rangle)|a_0\rangle = c_1|s_1\rangle|a_0\rangle + c_2|s_2\rangle|a_0\rangle \rightarrow c_1|s_1\rangle|a_1\rangle + c_2|s_2\rangle|a_2\rangle$$

In this derivation

$$|\psi\rangle_S = c_1|s_1\rangle + c_2|s_2\rangle$$

is state of quantum system(superposition), while

$$|\psi\rangle_{SA} = c_1|s_1\rangle|a_1\rangle + c_2|s_2\rangle|a_2\rangle$$

is state of combined system(entanglement/superposition).

The quantum measurement problem is the realization that the state

$$|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$$

is **not** observed as outcome of measurement!

What is seen is not a so-called superposition, but either

$$|s_1\rangle |a_1\rangle \text{ or } |s_2\rangle |a_2\rangle$$

That is the so-called **"problem of outcomes"**.

The state $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$

is usually referred to as a superposition, which is **misleading**.

Entanglement **outweighs** superposition as the **defining** feature of this state. Without entanglement-correlations - we would **not** have a measurement problem!

Careful investigation of this state in a 2007 experiment(Roch, et al) (a Wheeler-type delayed choice experiment) demonstrates its strikingly non-local character.

We will discuss the experiment in detail shortly.

A photon jumps from state $|\psi\rangle_S = c_1 |s_1\rangle + c_2 |s_2\rangle$ to state $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$

precisely when A(measuring device) switches on and while photon is still inside interferometer,

and jumps from $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ to $|\psi\rangle_S = c_1 |s_1\rangle + c_2 |s_2\rangle$ when A switches off during the experiment.

Quantum jumps, removal of interferences, and the observed non-locality are due to entanglement.

Of course, an entangled state = superposition, but it is a very **special** superposition.

To call $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ simply a "superposition" **misses** the crucial physics of entanglement and makes all the difference in the understanding of the state.

Entanglement is the characteristic trait of quantum mechanics; one that enforces the entire departure from classical lines of thought.

It is well known that, for 2-part systems, **all** non-product states exhibit non-locality (Bell inequality).

Thus when S and A are entangled $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$, they **share** a non-local **channel**.

The measurement state $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ is very subtle.

Although $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ is called a superposition of S and/or superposition of A, **neither** is true.

Thus, when S and A in measurement state $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$, neither S nor A in a superposition. $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ = superposition, but neither a superposition of S nor of A and also not superposition of states of composite system SA.

In $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$, S is in both states $|s_1\rangle$ and $|s_2\rangle$ simultaneously, as is known from observed interference between 2 states.

However, in case of $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ **experiment** now shows that SA is not in both two-part states $|s_1\rangle |a_1\rangle$ and $|s_2\rangle |a_2\rangle$ simultaneously, **but only simultaneously in two correlations between the states.**

The entanglement of **two** systems is quite **different** from superposition of **one** system.

Experiments(Rarity) demonstrated the **precise** sense in which $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ **represents** a superposition.

Experiments answer question "given that S and A are in state $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$, **what** (if any) entities interfere and **what** is nature of interference?"

Let us see exactly how……..

The Local State Solution(due to Jauch) of the Problem of Definite Outcomes

Consider a **single** quantum S(electron or photon), passing through double-slit experiment, with a "downstream" viewing screen.

Suppose an ideal "which-slit detector" A is present so that, upon detection, S and A become entangled in measurement state $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ with **orthogonal** detection states $|a_i\rangle$. Imagine S and A are separated by meters or kilometers.

Complete observation of experiment **requires** two "local observers", The 1st observer is S and the 2nd observer is A.

Such a non-local setup **has been carried** out experimentally.

What did the **local** observers observe this experiment?

A well-known prediction of quantum physics says

that the 1st observer observes implications of the **"local state of S,"**

represented by a **"reduced"** density operator,

where the degrees of freedom for the 2nd system

are averaged over(removed) by the "trace" (Tr) operation (see below)

$$\rho_S = \mathrm{Tr}_A(\rho_{SA}) = |c_1|^2 |s_1\rangle \langle s_1| + |c_2|^2 |s_2\rangle \langle s_2|$$

and 2nd observer would observe the implications of **"local state of A"**, represented by the reduced density operator

$$\rho_A = \mathrm{Tr}_S(\rho_{SA}) = |c_1|^2 |a_1\rangle \langle a_1| + |c_2|^2 |a_2\rangle \langle a_2|$$

where a density operator gives probabilities via the relation

$$P(b) = \mathrm{Tr}(\rho P_b) = \mathrm{Tr}(\rho |b\rangle \langle b|)$$

The "local state of S" is found by **completely removing** from the density operator any effects of A and vice versa.

This is the important idea!

Derivation of reduced density operator (for completeness)

$$\rho_{SA} = |\psi_{SA}\rangle\langle\psi_{SA}| = (c_1|s_1\rangle|a_1\rangle + c_2|s_2\rangle|a_2\rangle)(c_1^*\langle s_1|\langle a_1| + c_2^*\langle s_2|\langle a_2|)$$

$$\rho_S = \mathrm{Tr}_A(\rho_{SA}) = \sum_{k=1}^{2}\langle a_k|\rho_{SA}|a_k\rangle$$

$$= \sum_{k=1}^{2}\langle a_k|(|\psi\rangle_{SA\ SA}\langle\psi|)|a_k\rangle$$

$$= \sum_{k=1}^{2}\langle a_k|((c_1|s_1\rangle|a_1\rangle + c_2|s_2\rangle|a_2\rangle)(c_1^*\langle s_1|\langle a_1| + c_2^*\langle s_2|\langle a_2|))|a_k\rangle$$

$$= \sum_{k=1}^{2}(c_1|s_1\rangle\delta_{k1} + c_2|s_2\rangle\delta_{k2})(c_1^*\langle s_1|\delta_{k1} + c_2^*\langle s_2|\delta_{k2})$$

$$= |c_1|^2|s_1\rangle\langle s_1| + |c_2|^2|s_2\rangle\langle s_2|$$

where we have used orthonormality of measurement states via $\langle a_i \,|\, a_j\rangle = \delta_{ij}$

We note that the Tr (trace) operation = sum over all designated states

removes all knowledge of the designated(traced) system from the equation.

This is clearly useful i**f we do not know much** about the reduced system.

A similar result holds for ρ_A .

Now, continuing our discussion.

Clearly, the reduced states are **mixtures**, not superpositions.

QM predicts both LOCAL observers **find** mixtures **not** superpositions.

An ensemble of experimental trials verifies

this via mixed-state patterns in agreement with the assertion made earlier

that neither S nor A is in a superposition.

For a different example,

$$\rho_A = \mathrm{Tr}_S(\rho_{SA}) = |c_1|^2 |a_1\rangle \langle a_1| + |c_2|^2 |a_2\rangle \langle a_2|$$

predicts Schrödinger's cat is in a mixture of either dead or alive,

not a superposition of both dead and alive.

The **local states**

$$\rho_S = \mathrm{Tr}_A(\rho_{SA}) = |c_1|^2 |s_1\rangle \langle s_1| + |c_2|^2 |s_2\rangle \langle s_2|$$

and

$$\rho_A = \mathrm{Tr}_S(\rho_{SA}) = |c_1|^2 |a_1\rangle \langle a_1| + |c_2|^2 |a_2\rangle \langle a_2|$$

must be taken seriously as implying outcomes predicted to be observed at the two sites.

The local states **cannot be dismissed** simply

by the argument that the only “real” state is the “global state”

$$|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$$

There is no contradiction between the predicted local **mixtures**

$$\rho_S = \mathrm{Tr}_A(\rho_{SA}) = |c_1|^2 |s_1\rangle \langle s_1| + |c_2|^2 |s_2\rangle \langle s_2|$$

$$\rho_A = \mathrm{Tr}_S(\rho_{SA}) = |c_1|^2 |a_1\rangle \langle a_1| + |c_2|^2 |a_2\rangle \langle a_2|$$

and the unitarily-evolving global **pure state** .

$$|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$$

Important:

We are saying: To consider the combined system SA as a single system only evolving unitarily **misses** the essential physics of nonlocality.

Haroche and Raimond weighed in on the question of

when a two-part system should be considered as a composite of two subsystems,

versus when it should be considered a single system B = SA. .

The composite system should be considered as single whenever the binding between the parts is much stronger than the interactions involved in dynamics, so that the internal structure of the composite system is left unchanged as it travels through the experiment

By this criterion, SA not single system.

Not only does the relation between S and A change during experiment, the relation is entangled and thus nonlocal.

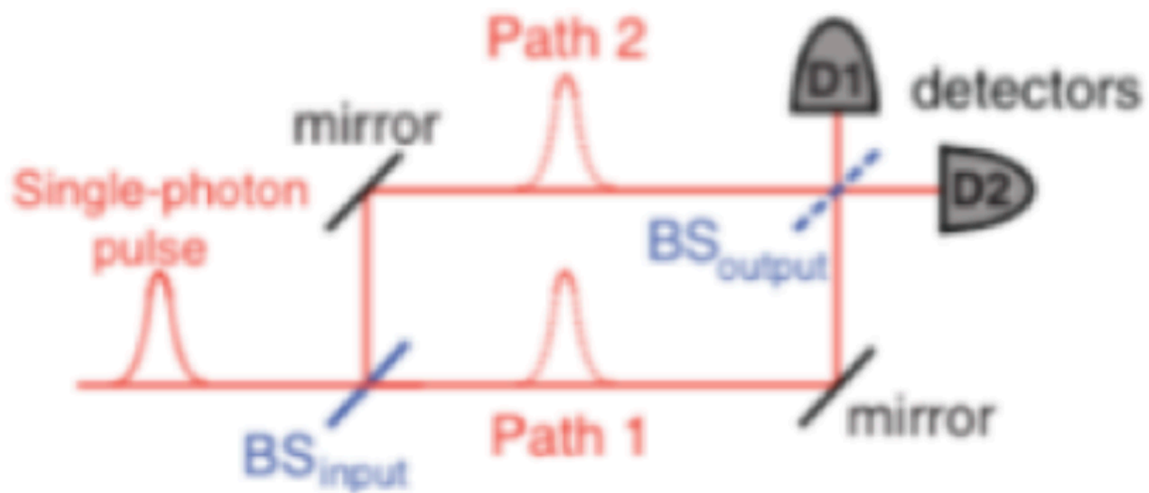
The implication is that $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ **must be** considered an entanglement of two separate systems S and A, not superposition of single composite system SA.

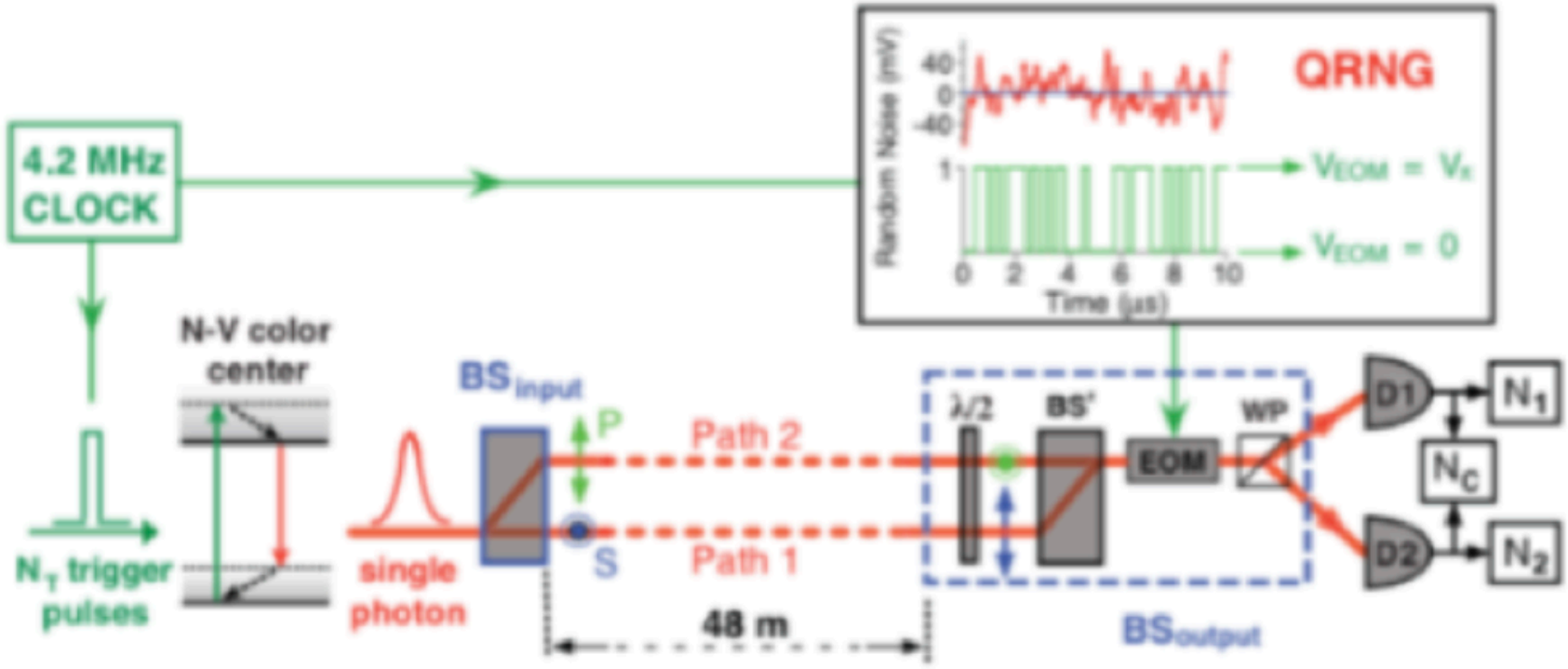
That's the basics of the theory at this point in my development! It follows completely from standard QM.

Not sure yet! Let us do it again!

Now for even more details and some possible new interpretations

Roch's "delayed choice" experiments used a Mach-Zehnder interferometer (1st figure) rather than the logically equivalent double-slit setup, to observe photons. It is shown in detail in the 2nd figure.





While a photon is on the **48-meter-long** interferometer paths, a quantum-based random number generator "decided" whether the second beam splitter (positioned at the end of paths - note the spatial separation) would be incorporated or omitted, i.e., it decided whether detectors would not or would (respectively) determine "which path."

In trials incorporating second beam splitter, an interference pattern is observed, indicating the photon passed through device as superposition along both paths.

On trials omitting second beam splitter, no interference is observed, indicating the photon passed through device along one or other path.

The two parallel paths were 5 millimeters apart.

Precisely (so far as the experiment could determine)

when the second beam-splitter switched from “on” to “off,” the photon collapsed in **mid-flight** from being on both paths to being on one or the other path.

The quantum jump was **correlated** with and **coincident** with

the incorporation or the omission of the second beam splitter

(i.e., with the **decision** to not entangle **or** to entangle the detectors with the photon).

All experimental results **are** just as predicted by the reduced states

$$\rho_S = \mathrm{Tr}_A(\rho_{SA}) = |c_1|^2 |s_1\rangle \langle s_1| + |c_2|^2 |s_2\rangle \langle s_2|$$

$$\rho_A = \mathrm{Tr}_S(\rho_{SA}) = |c_1|^2 |a_1\rangle \langle a_1| + |c_2|^2 |a_2\rangle \langle a_2|$$

Switching the second beam-splitter off entangles the photon and the detector in measurement state

$$|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$$

causing the photon to collapse from a superposition $|\psi\rangle_S = c_1 |s_1\rangle + c_2 |s_2\rangle$

into the mixture $\rho_S = \mathrm{Tr}_A(\rho_{SA}) = |c_1|^2 |s_1\rangle \langle s_1| + |c_2|^2 |s_2\rangle \langle s_2|$

that is observed at the detector.

At this point we are in the diagonal density matrix stage.

Thus, we are doing a classical measurement(like dice) and it should be interpreted as such!

This should resolve problem of definite outcomes!

This was my belief at this point in my tortuous path to understanding!

Quantum theory predicts and experiment verifies that, with the detector in operation,

observers of S and of A find them to be in definite mixtures,

not indefinite superpositions.

Any strategy of imagining the quantum and the detector to be widely separated

obviously changes nothing

- it does not matter whether the quantum and the detector are close together or far apart.

The key to understanding quantum measurements comes from

understanding nonlocal relationship that develops between S and A

when they evolve unitarily into entangled measurement state

$$|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$$

In experiments, this global state $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$

violates Bell's inequality, implying instantaneous non-local transfer of **correlations** across arbitrarily large distances.

Not information transfers, but
CORRELATIONS which we can only see later!

Without entanglement one would observe different results

- the interferences would not disappear as in the experiments.

But entanglement “decoheres-collapses” coherent states

so that S and A impact their detectors randomly.

Locally, entanglement “decoheres” each photon so that they exhibit definite outcomes.

But quantum dynamics is unitary, implying that the global state

$$|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$$

remains coherent despite the incoherence of subsystems.

Since the individual photons are now incoherently mixed, what has happened to coherence?

Answer:

It is the experimentally observed global coincidence measurements

which compare the impact points of entangled pairs.

In the double-slit experiment with two screens, each photon “knows” the impact point (i.e., phase shift) of other photon and instantly adjusts own impact point in order to form an interference pattern as a function of difference between two photons’ phase shifts!

This is strikingly non-local, and the experimental results violate Bell’s inequality.

Thus the coherence of the entangled state resides in correlations between subsystems, rather than in the subsystems themselves.

Entanglement **transforms** the coherence of states of S into

coherence of correlations **between** states of S and A, **allowing** S and A to exhibit definite outcomes while **preserving** the global coherence as demanded by unitary evolution.

We can now answer the question:

Precisely what is superposed and what interferes in the measurement state?

The answer is surprisingly simple:

Only the correlations between S and A are superposed.

Thus the measurement state $|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$ should be read as:

The state $|s_1\rangle$ is positively correlated with the state $|a_1\rangle$,
and the state $|s_2\rangle$ is positively correlated with the state $|a_2\rangle$

Only correlations are superposed, not states.

When the superposition $|\psi\rangle_S = c_1 |s_1\rangle + c_2 |s_2\rangle$ of S entangles with states of A, the superposition **shifts**, from a superposition of states of S to superposition of correlations between S and A, so S can be in an incoherent mixture **while** maintaining unitary global dynamics.

This is how nature resolves problem of definite outcomes.

So the coherence exhibited by the measurement state

$$|\psi\rangle_{SA} = c_1 |s_1\rangle |a_1\rangle + c_2 |s_2\rangle |a_2\rangle$$

must be **invisible** to local observers,

and **yet show up** in the global measurement state in order to **preserve** unitary dynamics.

One could regard this as the **underlying reason** why entanglement (i.e., measurement)

must shift coherence from the states of S and A to the correlations between S and A.

"Collapse" or the "appearance of definite values" (reduced density operator)

can be viewed as a consequence of the measurement state's nonlocality

plus special relativity's ban on instant signaling.

We note that the global measurement state is a very **different** animal from local states.

While the local mixed states are immediately observed at both local sites,

the global state can be "observed" only at some time after measurement by traveling to both local sites, gathering data from both, and then assembling information and noting the correlations between two sets of data.

Our conclusions so far:

When the detector measures the superposed quantum,

quantum physics predicts that the states actually observed are

local (i.e. mixed or reduced) states of subsystems,

it is not the superposed global state that follows from Schrödinger's equation.

The local states directly observed in the measurement

must contain **no** hint of nonlocal correlations

between two subsystems since this would violate relativity's prohibition on instant signaling.

Thus the local states describe what **actually** happens at both subsystems.

The global state predicts these local states(they can be derived from it),

and **also** predicts indirectly-observable (by gathering global data at **later** time)

nonlocal correlations between these states.

Continuing along my path to understanding…..

An even more dramatic experiment — Experimental nonlocality and entanglement

Now we demonstrate that, according to standard quantum theory (**and** experiment),

measurement state (5) represents none of paradoxical situations we have mentioned.

Again let us repeat material from earlier - understanding this topic is so important that repetition is warranted.

The **unity** of the quantum suggests that the measurement state (5) represents a unified,

hence superposed and pure, quantum state of composite system.

We asked the question: precisely what is superposed?

Studied simple (i.e. non-composite) superposition (3) via interference exhibited in M-Z experiment.

Varying the length of either path 1 or path 2

created varying interference effects in detectors,

demonstrating each photon really must travel both paths to its detector.

Quantum theory agrees entirely with these conclusions,

as can be shown by using photon wavelengths

to show that path differences correctly predict interferences observed at each detector.

This implies that to understand the measurement state,

one needs to find and analyze entanglement experiments that demonstrate interference.

This has been done for several decades in connection with quantum non-locality.

Many nonlocal interference experiments have been done

beginning with Clauser and Freedman,

culminating in experiments demonstrating nonlocality across great distances

and that simultaneously closed all possible loopholes in all previous experiments.

By now, it is well known that entangled state (5) predicts nonlocal effects

between two subsystems, and that phase variations of either subsystem

cause instantaneous(non-signaling), i.e., non-local, readjustments(correlations)

of the possibly-distant other subsystem.

When macroscopic systems are involved (i.e., cats) we have a problem.

it is not easy to **vary the phase of cat**,

and as we saw in the Mach-Zehnder experiment,

one **cannot** understand superposition without varying phases of the superposed parts.

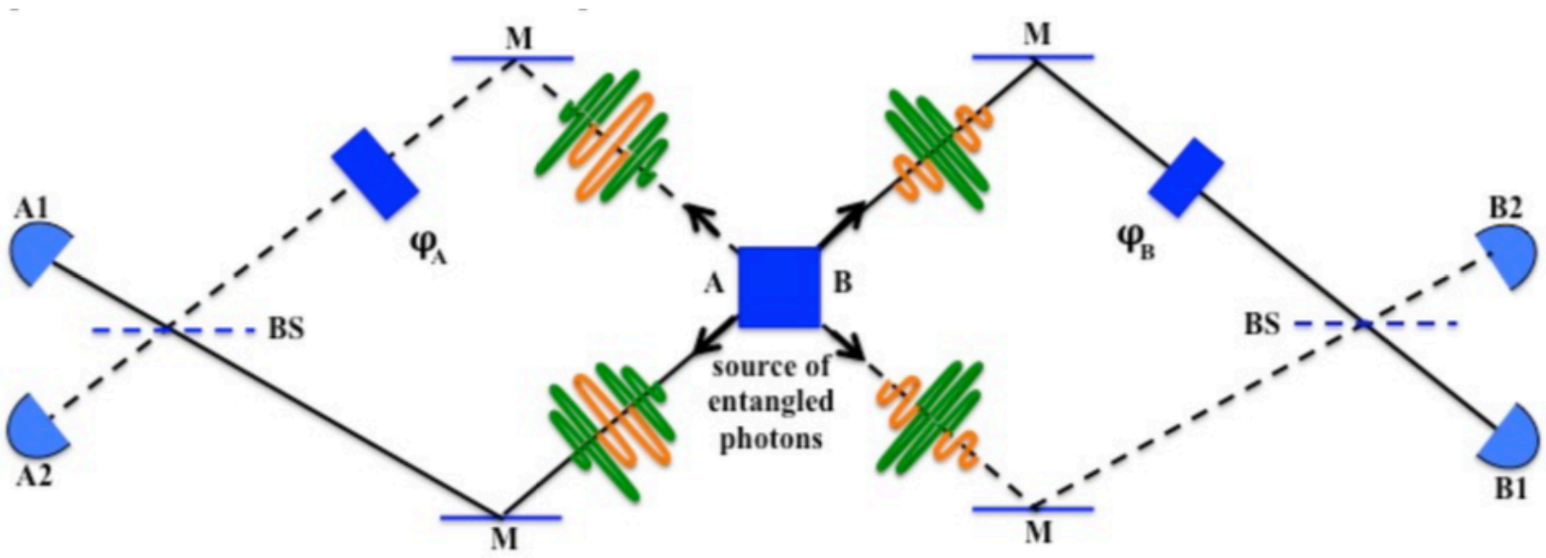
Thus, all nonlocality experiments carried out with pairs of simpler quanta such as photons.

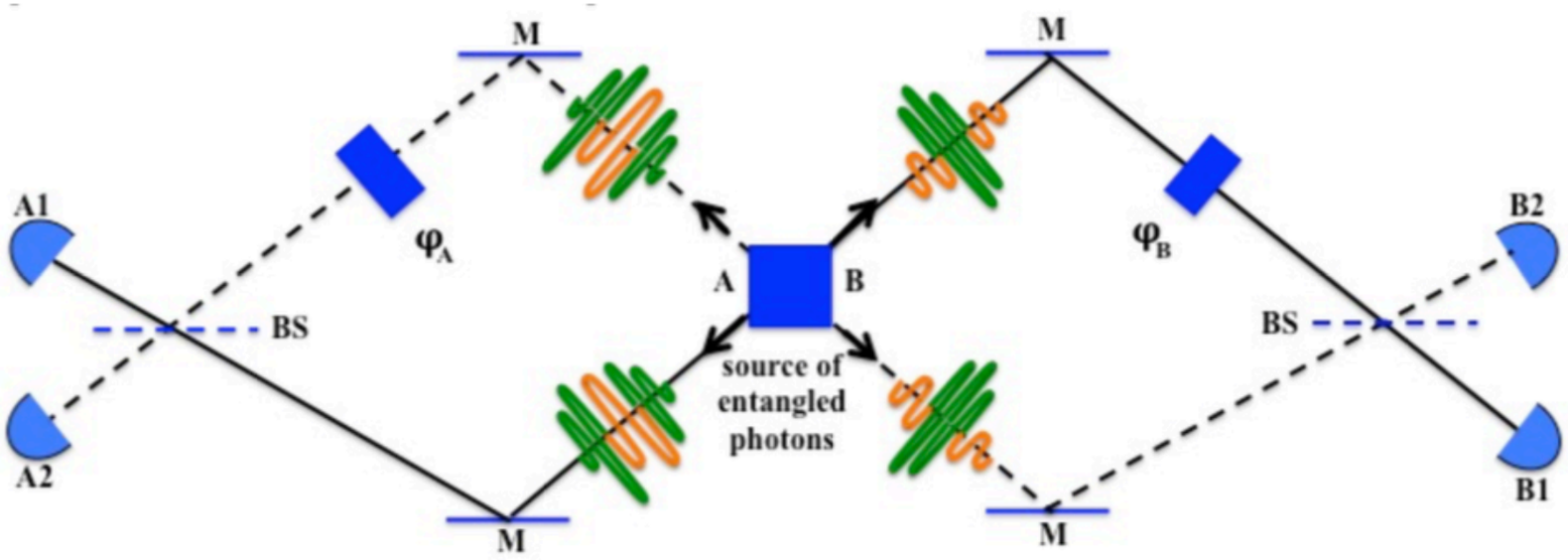
The best nonlocal entanglement experiments

most appropriate for investigating measurement

were conducted nearly simultaneously by Rarity,Tapster and Ou, Zou, Wang (RTO).

Figure shows layout for these "RTO" experiments.





The “Source" creates entangled photon pairs by "parametric down-conversion”, a process which we discussed earlier.

The RTO experiment is **two back-to-back MZ-interferometer experiments** but with first beam splitter for each photon located **inside** the source of entangled photons.

Without entanglement, each single photon (either A or B) would interfere with itself at its own detectors according to own phase shifters ϕ_A or ϕ_B.

In the experiment, two entangled photons are emitted into a superposition of solid paths connecting detectors A1 and B1, and dashed paths connecting detectors A2 and B2.

Remember that the two photons are already entangled when emitted.

Entanglement changes everything.

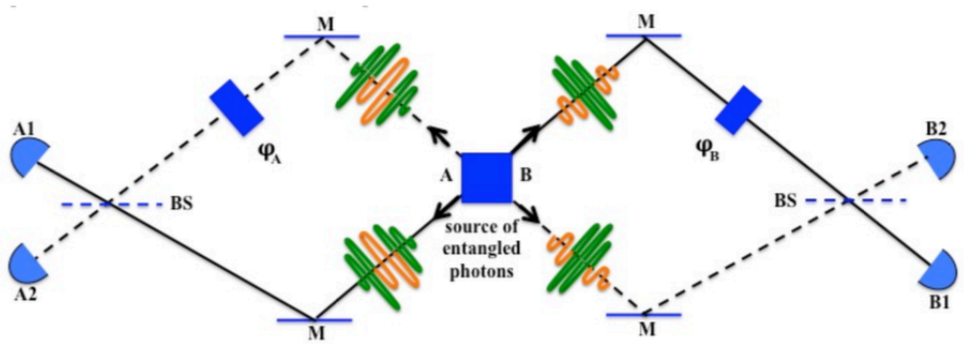
No longer does either photon interfere with itself at own detectors.

Instead, photons are entangled in measurement state (5) with $|\psi_1\rangle$ and $|\psi_2\rangle$

representing (say) solid-line and dashed-line states of A and $|1\rangle$ and $|2\rangle$

representing solid-line and dashed-line states of B,

although in RTO experiments **neither** subsystem is macroscopic.



Each photon **now** acts like a which-path detector for the other photon.

Recall the double-slit experiment:

When the which-slit detector is switched on, the pattern on screen switches

abruptly from a striped interference pattern indicating the pure state nature

of each electron across both slits, to phase-independent sum

of two non-interfering single-slit patterns.

Entanglement between the electron and the which-slit detector breaks the pure state into two single-slit parts, so that measured electron comes through either slit 1 or slit 2.

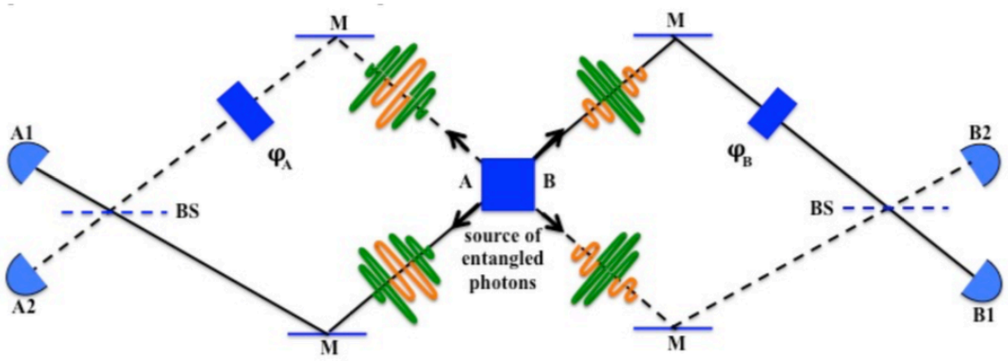
This suggests that in the RTO experiment, the entanglement should break the pure-state superposition (3) into two non-interfering parts.

This is exactly what is observed.

Both photons impact their detectors as random 50-50 mixtures, just like a flipped coin.

Entanglement **breaks** the single-photon pure state (3) observed in the Mach-Zehnder experiment, causing each photon to behave "incoherently" with no dependence on its phase setting.

But (5) is pure state.



Where has the phase dependence gone?

The **answer** lies in the phase-dependent but nonlocal relationship observed **between** the solid and the dashed branches.

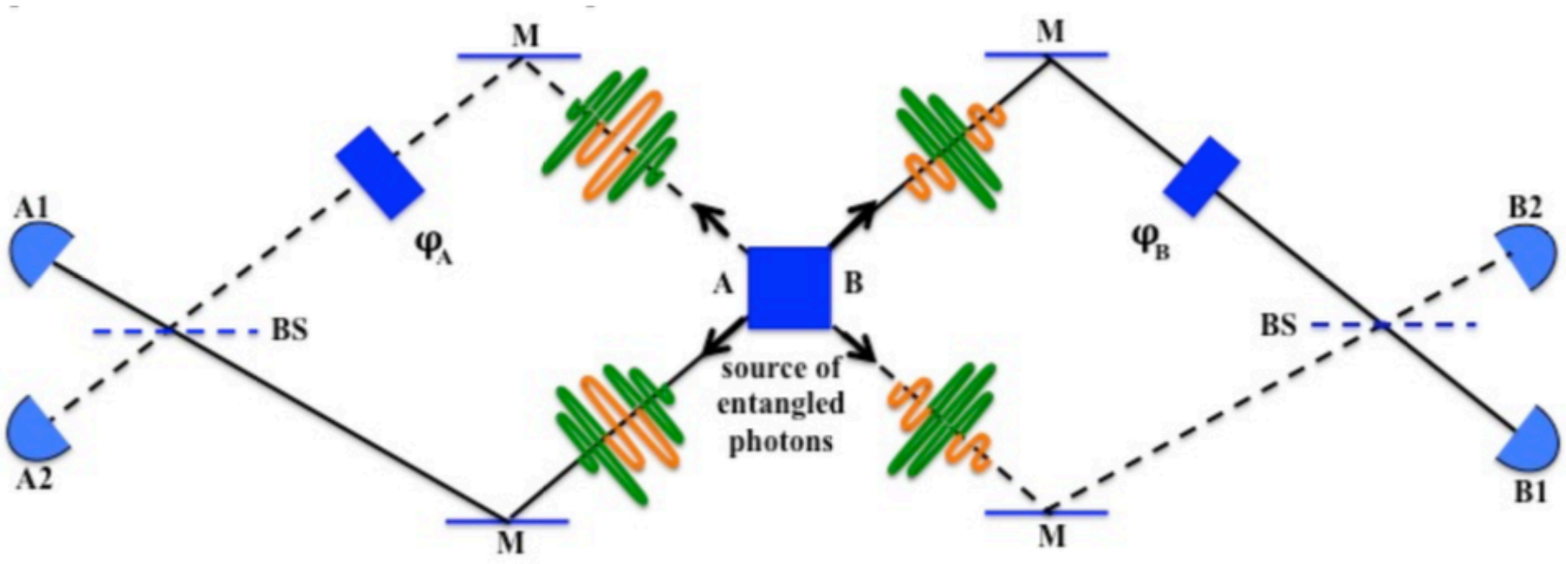
This phase dependence is observed experimentally in coincidence (or **correlation**) measurements comparing the detections of entangled pairs.

The “flipped coins” mentioned above turn out to be correlated with each other.

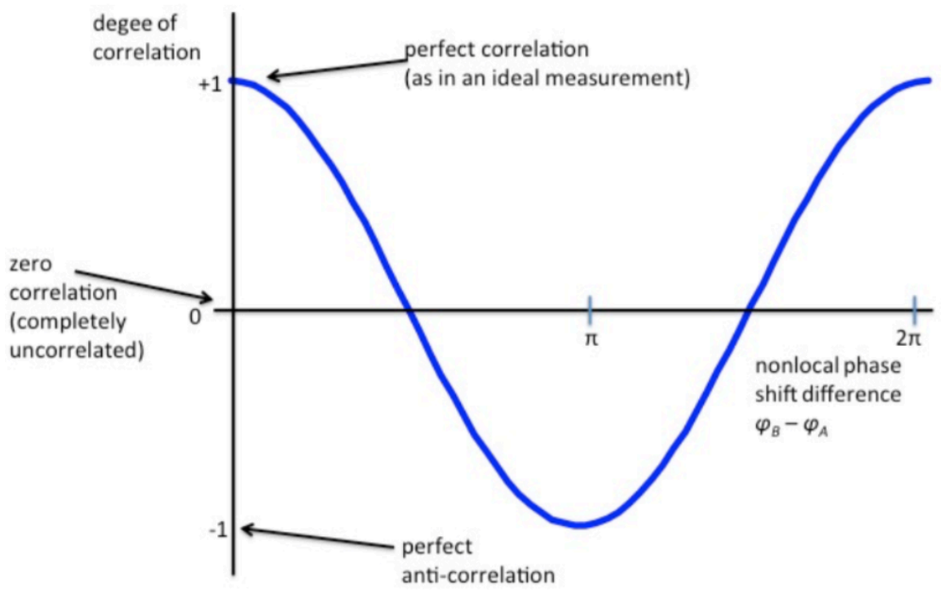
This phase dependence across two separated subsystems is **essential to preserve** unity of (now entangled) quantum.

This is **not** an easy experiment to perform:

The source creates a stream of photon pairs, and one must compare the impact of single photon A at detectors A1, A2 with impact of corresponding entangled photon B at detectors B1, B2.



RTO figured out how to do this, with result shown in Figure.



The figure graphs the **degree of correlation** between A and B.

This is a **measure of the agreement** between the outcomes at A's detectors and B's detectors.

A correlation of +1 means perfect, or 100%, agreement:

Either both sets of detectors register outcome 1 (i.e., A1 and B1 click)

or both register outcome 2.

The opposite extreme is a correlation of -1, meaning 100% disagreement: If one detector registers 1, the other registers 2.

Either correlation, +1 or -1, implies that either photon's outcome is **predictable** from the other photon's outcome.

A correlation of zero means one photon's outcome does not at all determine other's outcome: Each photon has random 50-50 chance of either outcome regardless of the other photon.

Correlations between 0 and +1 mean the outcomes are more likely to agree than to disagree, with larger correlations denoting higher probability of agreement; for example, correlation of +0.5 means 75% probability of agreement.

Similarly, correlations between 0 and -1 mean outcomes are more likely to disagree than to agree; a correlation of -0.5 means a 75% probability of disagreement.

The RTO experiment agrees entirely with the predictions of standard quantum physics.

When accounting is made of optical paths for both photons, QM obtains the following result:

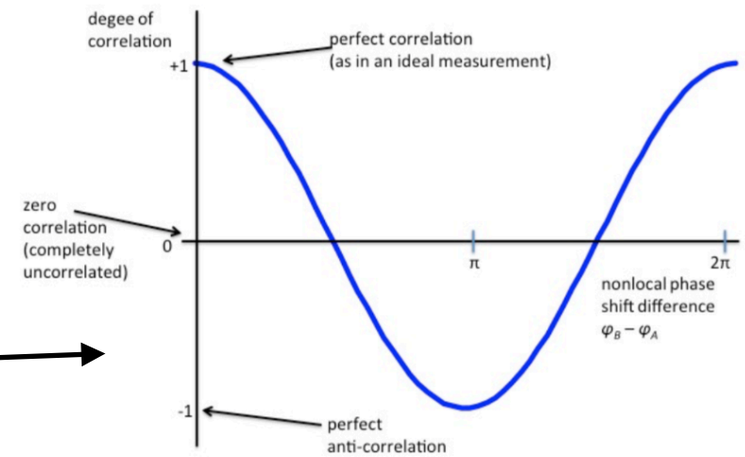
$$P(correlated) = P(A1\ and\ B1) + P(A2\ and\ B2) = \frac{1}{2}[1 + \cos(\varphi_B - \varphi_A)]$$

$$P(anticorrelated) = P(A1\ and\ B2) + P(A2\ and\ B1) = \frac{1}{2}[1 - \cos(\varphi_B - \varphi_A)]$$

where P(correlated) is single-trial probability that A's and B's detectors will agree,

and P(anticorrelated) is a single-trial probability that A's and B's detectors will disagree.

The degree of correlation, defined as P(correlated) – P(anticorrelated), is then simply $\cos(\phi_B - \phi_A)$, as graphed in last figure.



In 1964, John Bell, as we discussed earlier, published a ground-breaking article stating the sufficient condition for a statistical theory such as quantum physics to meet a condition known as "locality."

He defined locality to mean **"that the result of measurement on one system be unaffected by operations on distant system with which it has interacted in past"**.

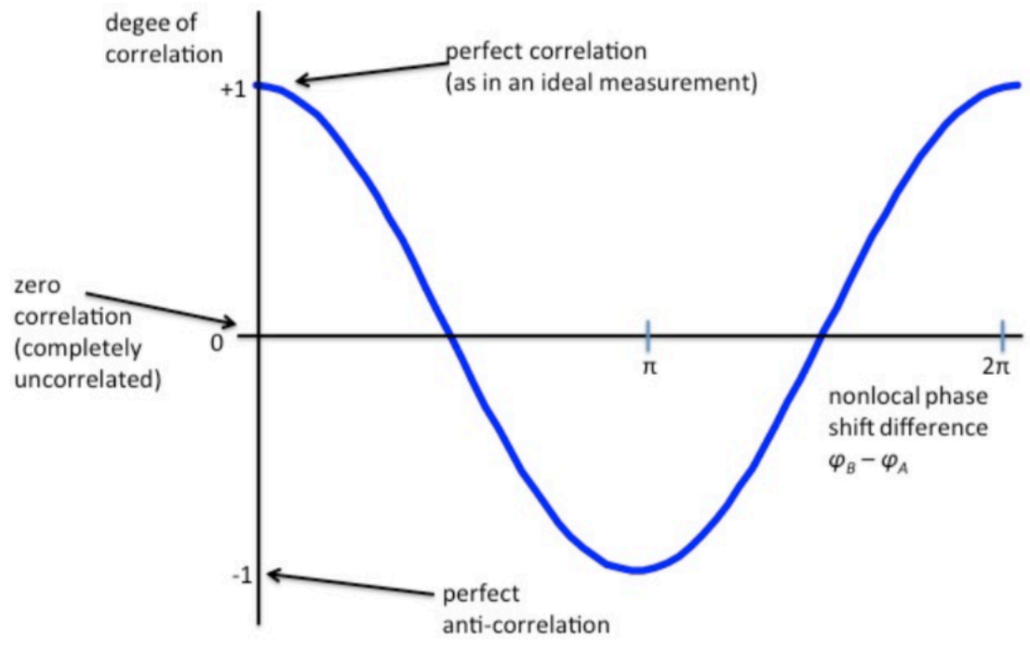
Bell expressed a sufficient condition in the form of inequality that any local theory must obey.

He demonstrated that certain statistical predictions of quantum physics violate Bell's inequality, i.e. quantum physics makes nonlocal predictions.

The results in last figure implies the case in point: the figure violates Bell's inequality at all phase differences $\phi_B - \phi_A$ other than 0, π, and 2π.

Let me **<u>underline</u>** meaning of this:

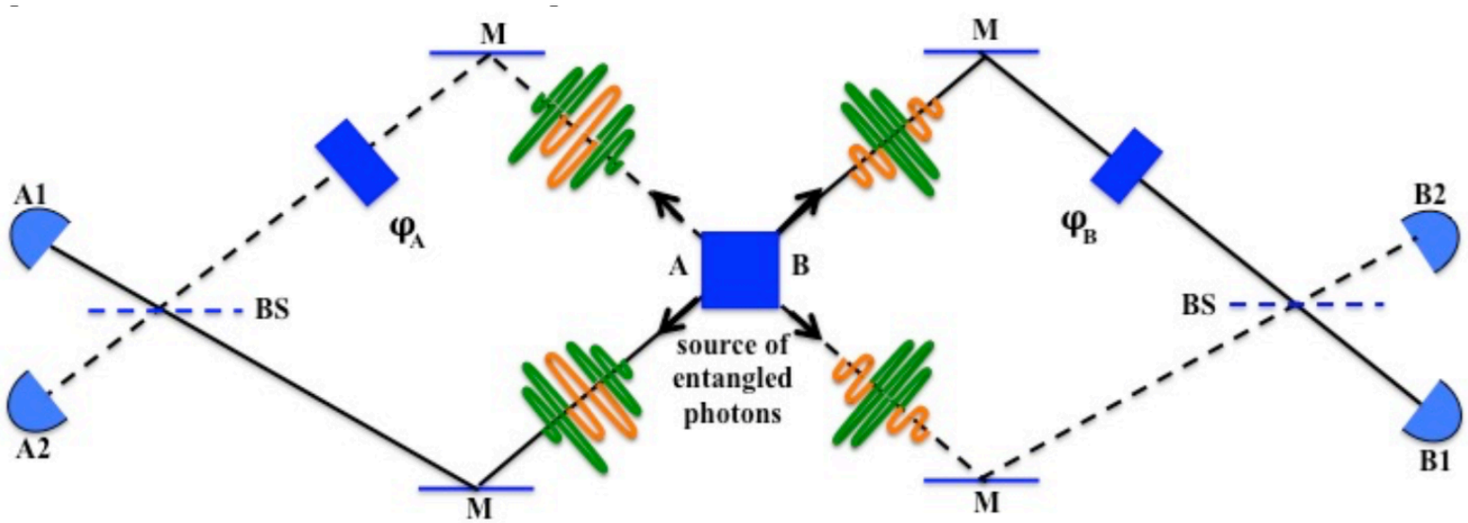
Violation of Bell's inequality means that the statistics of the measurements on photon A - photon A's **"statistical behavior"** - is necessarily affected by the setting of photon B's phase shifter.



In fact, even without Bell's condition, the nonlocality of the experiment is intuitively obvious.

Here's why:

Suppose we set the phase shifters to zero and that all four optical paths (two solid, two dashed) in are then equal; thus $\phi_B - \phi_A$ is zero.



Without the two beam splitters BS,

the two photons emitted into the solid pair and the dashed pairs of paths

would impact either detectors A1 and B1 or A2 and B2

because of the symmetry of experiment and conservation of momentum.

This is neither surprising nor nonlocal,

and would happen even if the photons were not entangled.

But the beam splitter is a **randomizing** device

that mixes the solid and dashed paths;

—> any photon passing through it has 50-50 chance of reflection or transmission.

With non-entangled photons and both beams splitters in place, there would then be no correlation between photon A's outcome and B's outcome because the two photons are independent of each other.

With entanglement, the correlation is perfect.

How does one photon know which path the other photon took at the other photon's beam splitter? **poor question!!**

Each photon is now "detecting" the quantum state of the other photon, from a distance that could be large.

Perfect correlation certainly "feels" nonlocal even though (as mentioned above) this perfect correlation at $\phi_B - \phi_A = 0$ does not violate Bell's inequality.

Note that such a violation is a sufficient but not a necessary condition for nonlocality.

Non-locality is **written** all over the RTO experiment.

Each photon "knows" which direction the other photon takes at its beam splitter and adjusts its selection accordingly.

The key nonlocal feature of graph, which is simply a cosine function, has $(\phi_B - \phi_A)$ as its independent variable.

Thus any desired shift in correlations can be made by an observer at either of
possibly-widely-separated phase shifters.

Bell suspected that this kind of situation meant that observer A (call her Alice)
could use her phase shifter to alter outcomes that would have occurred at both her own and observer B's (call him Bob) detector and, following up on this hypothesis, derived his inequality involving probabilities at both Alice's and Bob's detectors
which, **if violated**, implied that both photons must have readjusted their states.

Such a readjustment is just what we expect,
given the **unity** of the quantum
and thus the unity of atoms and other entangled systems such as our two photons.

The two photons form single **"bi-quantum", an "atom of light"**, in the pure state (5).

When Alice varies her phase shifter, both photons "know" both path lengths and **readjust** their behavior accordingly to produce the proper correlations.

Analogously, a single photon "knows" both path lengths in single-photon interferometer experiment.

Finally, we come to the central question of the discussion:

What is actually superposed in the entangled superposition (5)?

A Mach-Zehnder experiment tests the simple superposition (3), while the RTO experiment tests the entangled superposition (5).

We know what is superposed in Mach-Zehnder, namely quantum states $|\psi_1\rangle$ (path 1) and $|\psi_2\rangle$ (path 2). This is deduced from the effect that either phase shifter has on both states.

Now consider the RTO experiment.

What is the effect of shifting either phase shifter?

One thing that does not change is the state ("local state" would be a better term, as discussed earlier) of either photon A or photon B:

As we know, both photons remain in 50-50 mixtures regardless of either phase setting.

What does change with variations in either phase shifter is the correlations between A and B.

With $\phi_B - \phi_A = 0$ have perfect correlation:

Either A1 and B1 (which we denote (11)) or A2 and B2 (denoted (22)).

As we vary either ϕ_B or ϕ_A we obtain non-zero probabilities of anti-correlated individual trials,

denoted (12) (outcomes A1 and B2) and (21) (A2 and B1).

When the non-local phase angle difference ($\phi_B - \phi_A$) reaches π/2,

we have zero correlation, and when it reaches π we have perfect anti-correlation.

Table 1 summarizes the crucial points in more detail.

Simple superposition:		**Entangled superposition of two sub-systems:**		
φ	State of photon	$\varphi_B - \varphi_A$	State of each photon	Correlation between the two photons
0	100% "1", 0% "2"	0	50-50 "1" or "2"	100% corr, 0% anti
π/4	71% "1", 29% "2"	π/4	50-50 "1" or "2"	71% corr, 29% anti
π/2	50% "1", 50% "2"	π/2	50-50 "1" or "2"	50% corr, 50% anti
3π/4	29% "1", 71% "2"	3π/4	50-50 "1" or "2"	29% corr, 71% anti
π	0% "1", 100% "2"	π	50-50 "1" or "2"	0% corr, 100% anti

Table 1. In a simple superposition, the photon's state varies with phase angle. In an entangled superposition, the relationship between states of the two photons varies, while individual states of both photons are phase-independent (or "mixed").

The column titled "simple superposition" shows how the superposition state of single photon

(M-Z) varies from "100% state 1" to "100% state 2"

as phase angle between two states varies.

Column titled "entangled superposition of two subsystems" shows

that the state of each photon remains unchanged throughout entire range

of both phase settings, while nonlocal correlation between the states of two photons

varies from "100% correlated" to "zero correlation" and then to "100% anticorrelated"

as either of the two local phase angles varies.

So, once again, what is superposed in the RTO experiment?

The hallmark of a superposition is the dependence on phase difference between the objects superposed.

But Table 1 exhibits **no** such phase dependence of the states of the two photons.

Each photon remains in an unchanging 50-50 mixture of own "path 1" and "path 2" states

- a situation that is radically at odds with the true superposition of path 1 and path 2 exhibited by the M-Z experiment.

Thus, in the entangled RTO state, **neither** photon is superposed.

We see here the source of the "classical" or non-superposed
nature of reduced density operators (Eqs. (10) and (11)),
not to mention the non-superposed
and hence non-paradoxical nature of Schrodinger's cat.

Simple superposition:		**Entangled superposition of two sub-systems:**		
φ	State of photon	$\varphi_B - \varphi_A$	State of each photon	Correlation between the two photons
0	100% "1", 0% "2"	0	50-50 "1" or "2"	100% corr, 0% anti
$\pi/4$	71% "1", 29% "2"	$\pi/4$	50-50 "1" or "2"	71% corr, 29% anti
$\pi/2$	50% "1", 50% "2"	$\pi/2$	50-50 "1" or "2"	50% corr, 50% anti
$3\pi/4$	29% "1", 71% "2"	$3\pi/4$	50-50 "1" or "2"	29% corr, 71% anti
π	0% "1", 100% "2"	π	50-50 "1" or "2"	0% corr, 100% anti

Table 1. In a simple superposition, the photon's state varies with phase angle. In an entangled superposition, the relationship between states of the two photons varies, while individual states of both photons are phase-independent (or "mixed").

Examination of the phase-dependence of the measurement state (5), as demonstrated by nonlocality experiments such as the RTO experiment,

reveals the true nature of Schrodinger's cat.

The last column of Table 1 shows us what actually is superposed when two subsystems are entangled in measurement state (5).

Since the correlations between two photons vary sinusoidally as the non-local phase angle between the two photons varies, clearly these are correlations between the states of two photons, and not the states themselves, that are interfering.

Entanglement has shifted superposition, from states of one photon A (Eq. (3), M-Z) to correlations between photon A and photon B (Eq. (5), RTO).

Simple superposition:		**Entangled superposition of two sub-systems:**		
φ	State of photon	$\varphi_B - \varphi_A$	State of each photon	Correlation between the two photons
0	100% "1", 0% "2"	0	50-50 "1" or "2"	100% corr, 0% anti
$\pi/4$	71% "1", 29% "2"	$\pi/4$	50-50 "1" or "2"	71% corr, 29% anti
$\pi/2$	50% "1", 50% "2"	$\pi/2$	50-50 "1" or "2"	50% corr, 50% anti
$3\pi/4$	29% "1", 71% "2"	$3\pi/4$	50-50 "1" or "2"	29% corr, 71% anti
π	0% "1", 100% "2"	π	50-50 "1" or "2"	0% corr, 100% anti

Table 1. In a simple superposition, the photon's state varies with phase angle. In an entangled superposition, the relationship between states of the two photons varies, while individual states of both photons are phase-independent (or "mixed").

More learning, More repetition, More Details and Some Tentative Conclusions

In order to resolve the problem of definite outcomes of measurements, aka Schrodinger's cat,

our discussion analyzed the entangled state (5) of a microscopic quantum

and its macroscopic measuring apparatus.

This state is a superposition of two composite entities $|\psi_1\rangle|1\rangle$ and $|\psi_2\rangle|2\rangle$,

with a phase angle between these entities that can range over 2π radians.

To resolve problem of definite outcomes we must ask:

Precisely what does the composite superposition (5) actually superpose, physically?

In order to understand a simple non-composite superposition (3),

we looked at the effect of varying the phase angle between

superposed entities $|\psi_1\rangle$ and $|\psi_2\rangle$

in an experimental setting such as M-Z interferometer.

The theoretically predicted and experimentally observed results then made it
obvious that the quantum whose state is (3) flows
simultaneously along two separate paths described by $|\psi_1\rangle$ and $|\psi_2\rangle$.

To understand superposition (5), one should proceed similarly by studying situations
in which the phase angle between superposed entities $|\psi_1\rangle|1\rangle$ and $|\psi_2\rangle|2\rangle$ varies.

Theorists and experimentalists studying the phenomenon of nonlocality
have been doing this for decades, but quantum foundations specialists have not particularly
noticed how the work is connected with the measurement problem.

In fact, the nonlocal aspects of the state (5) have been studied
since Bell's 1964 theoretical paper and Clauser and Freedman's 1972 experiment.
The 1990 experiments of Rarity and Tapster, and of Ou, Zou, Wang, and Mandel,
furnish an ideal vehicle for such an analysis and are the central feature of this discussion

One lesson of this analysis is that, in order to understand the measurement problem, one must understand the significance of nonlocality.

This is because the key measurement state (5) that caused Schrodinger and other experts so much concern has nonlocal characteristics.

It must be understood as a superposition of correlations, rather than a superposition of states, but this cannot become apparent until one considers the effect of variations in phase angle between its superposed terms.

Such variations are not part of the measurement process itself because measurements are designed to take place at zero phase angle.

Experimental or theoretical studies of such phase variations will have nonlocal ramifications, because such variations are inherently nonlocal!

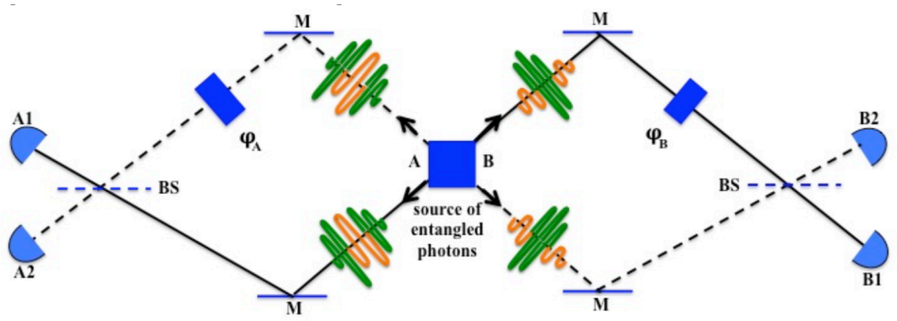
This situation would have prevented Schrodinger in 1935, or indeed anyone prior to Bell's 1964 paper and the experimental confirmations of reality of nonlocality beginning in 1972, from understanding entangled superposition (5).

It's worth emphasizing that, when two subsystems are entangled in measurement state (5), neither subsystem is superposed.

Only the correlations between subsystems are superposed.

In RTO experiments, the two correlations in question are represented by solid and dashed paths connecting pairs of outcomes.

A pair of photons entangled in state (5) follows both of these paths simultaneously.



The subsystems themselves, however, are not in superpositions, but instead, in indeterminate mixtures of definite states.

Thus observers of either subsystem will observe only definite outcomes, as predicted by local mixtures (10) and (11).

RTO experiments are entangled analog of the M-Z interferometer experiment: a pair of back-to-back interferometer experiments, with entangled pair of quanta of which one quantum passes through each interferometer.

As we said earlier, the experiment and its theoretical analysis shows that, when superposed photon A becomes entangled with second photon B to form state (5), the nonlocal aspect of A's superposition is transferred to correlations between A and B.

Thus, an entangled state such as (5) is neither a superposition of states of A nor of states of B, but is instead a superposition of correlations between states of A and states of B.

To see most clearly, compare simple superposition (3) with entangled superposition (5).

In the simple superposition, the state observed by a "which- state" detector varies smoothly from 100% $|\psi_1\rangle$, through 50% $|\psi_1\rangle$ and 50% $|\psi_2\rangle$, and finally to 100% $|\psi_2\rangle$ as the phase angle ϕ between $|\psi_1\rangle$ and $|\psi_2\rangle$ varies from 0 to π.

In entangled superposition, neither the state of A nor the state of B varies as ϕ_A or ϕ_B varies;

both A and B remain in 50-50 mixtures throughout.

What does vary is the correlation between A and B.

A non-local "correlation detector" (i.e., an RTO-type of experiment!)

would find the relation between the two subsystems varies from 100% positively correlated

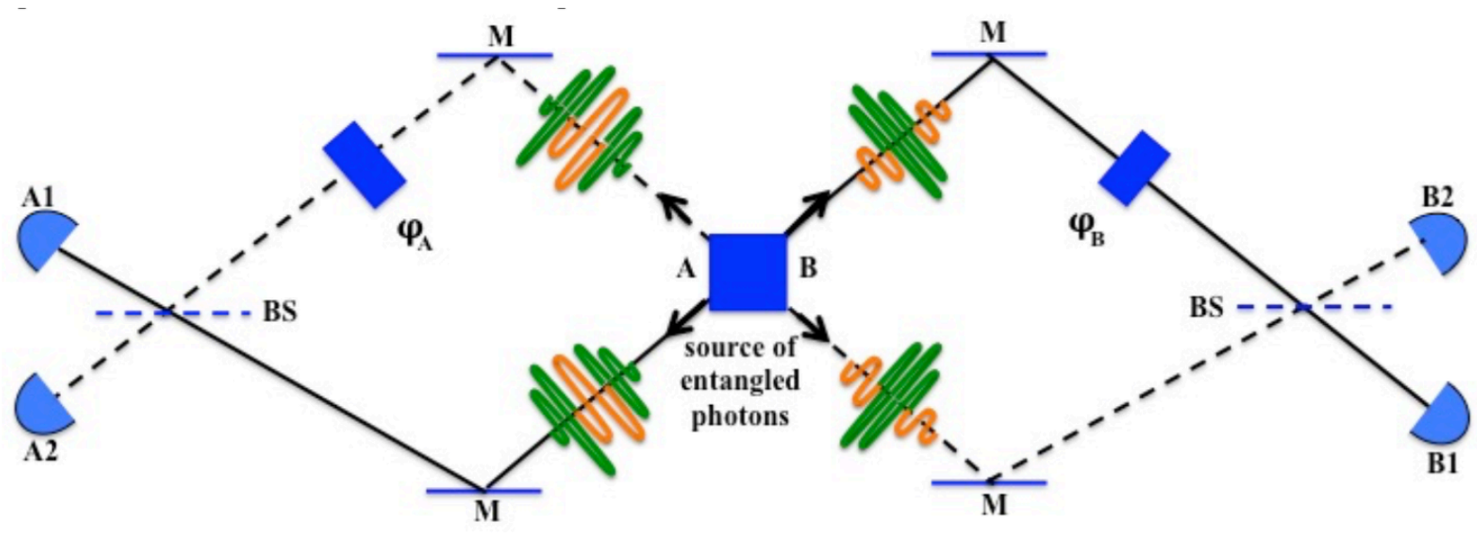
(either pair state 11 or 22, pictured by solid and dashed paths in figure),

to 50% positively correlated and 50% anti-correlated, and finally to 100% anti-correlated

(12 or 21), as the nonlocal phase difference $\phi_B - \phi_A$ varies from 0 to π.

This is a superposition of correlations,

not a superposition of composite states or of non-composite (single-system) states.



At least in idealized case of a minimally-disturbing von Neumann measurement, the initial stage of the measurement process (through formation of the measurement state (5)) can be **described** as follows:

A quantum in a simple superposition such as (3) entangles with a macroscopic which-path detector.

At the instant of entanglement, the local states of both the quantum and the detector undergo a radical change, **a quantum jump**.

Locally, the detector and the quantum jump into mixtures (10) and (11).

Simultaneously, the global state (5) continues evolving smoothly according to the Schrodinger equation.

Entanglement causes the superposed single quantum to be **instantly transformed** into superposed correlations between the quantum and the detector.

This stage of measurement process is **entirely describable** in terms of pure global states following Schrodinger equation.

"Collapse" from the local superposition to the local mixtures occurs because of the formation of the entangled state (5) **and** the resulting formation of subsystems whose local states (Eqs. (10) and (11)) have definite outcomes.

Note that the phenomenon of nonlocality is essential to preserving the pure-state nature (the unity) of the composite system.

To put this more intuitively, a reorganization throughout the entire extent of the composite entangled system is required in order to preserve unity of the (now entangled) quantum.

According to Table 1, when

two systems entangle to form the state (5),

both “collapse” into phase-independent

local mixtures.

Relativity **requires** this phase independence:

Simple superposition:		**Entangled superposition of two sub-systems:**		
φ	State of photon	$\varphi_B - \varphi_A$	State of each photon	Correlation between the two photons
0	100% “1”, 0% “2”	0	50-50 “1” or “2”	100% corr, 0% anti
$\pi/4$	71% “1”, 29% “2”	$\pi/4$	50-50 “1” or “2”	71% corr, 29% anti
$\pi/2$	50% “1”, 50% “2”	$\pi/2$	50-50 “1” or “2”	50% corr, 50% anti
$3\pi/4$	29% “1”, 71% “2”	$3\pi/4$	50-50 “1” or “2”	29% corr, 71% anti
π	0% “1”, 100% “2”	π	50-50 “1” or “2”	0% corr, 100% anti

Table 1. In a simple superposition, the photon's state varies with phase angle. In an entangled superposition, the relationship between states of the two photons varies, while individual states of both photons are phase-independent (or "mixed").

If any phase-dependent aspect of entangled state were locally observable, instant

information-containing messages could be sent, violating special relativity.

Local states of entangled subsystems must be invariant to phase changes.

Thus, only the relationship - the correlations -

between A and B, but not A or B themselves, can vary with phase angle.

Since local observers cannot detect these correlations,

the entangled state cannot be used to send superluminal signals.

This is, ultimately, the reason Schrodinger’s cat

must be either alive or dead rather than a superposition of both.

A phase-dependent superposition involving both local states

would permit nonlocal signaling, violating relativity.

This conclusion implies that the standard physical description of composite non-entangled (i.e. factorable) product state such as $|\psi 1\rangle|1\rangle$ **has, for a long time, been mistaken**.

Usually we regard $|\psi 1\rangle|1\rangle$ as state of the composite system AB, where subsystem A is in state $|\psi_1\rangle$ and subsystem B is in state I1>.

But this leads us into the **paradox** of Schrodinger's cat, where (5)

$$(|\psi_1\rangle\,|1\rangle + |\psi_2\rangle\,|2\rangle)/\sqrt{2}$$

represents a state in which two macroscopically different composite states exist **simultaneously** as a superposition.

According to the present discussion, quantum theory and quantum experiments **imply** this entangled state to be a **superposition** of correlations between states rather than a superposition of composite states.

Thus $|\psi_1\rangle|1\rangle$ is **not** a state of composite system, but **instead** a correlation between two subsystems.

That is, $|\psi_1\rangle|1\rangle$ means **"subsystem A is in state $|\psi_1\rangle$ if and only if subsystem B is in state $|1\rangle$,"** an important departure from usual description.

Even if one of two subsystems happens to be macroscopic detector, entangled state (5) is simply a non-paradoxical superposition of correlations.

It says merely that state $|\psi_1\rangle$ of A is correlated with state $|1\rangle$ of B, and state $|\psi_2\rangle$ of A is correlated with the state $|2\rangle$ of B, with **non-local phase angle** $\phi_B - \phi_A$ determining degree of each correlation.

Regardless of the value of the phase angle, neither subsystem is in a superposition.

The entangled measurement state (5) is best described as a "macroscopic correlation":

a pair of superposed (i.e. phase-dependent) quantum correlations in which one subsystem happens to be macroscopic.

It is technically very difficult to create a macroscopic superposition, but macroscopic which-path detectors routinely achieve state (5).

It's not paradoxical, even though many physicists have puzzled over it.

In entanglement, nature employs an ingenious tactic.

She must not violate relativistic causality, yet she must be nonlocal in order to maintain the pure-state nature of the original single-quantum superposition over composite objects such as bi-photons.

Thus she accomplishes nonlocality entirely via the superposition of correlations, because correlations cannot be locally detected and thus their superposition cannot violate relativity.

This tactic lies behind the nonlocal spread of phase-dependence over large spatial distances.

By means of the superposition of correlations - entanglement - nature creates a phase-dependent pure-state quantum structure across extended quantum systems such as bi-photons.

I've frequently used the term "local" as contrasted with "global."

For composite systems, and especially the entangled measurement state, it's a crucial distinction.

Entangled states such as (5) have distinct local and global (nonlocal) aspects.

The local description corresponds to two observers, each observing only one subsystem.

In the case of (5), this "local description" is fully captured by the reduced density operators (10) and (11) - **each local observer detects a mixture, not a superposition, of one subsystem.**

The "global description" means the evolving pure state of the entire composite system, in our case Eq. (5).

It is a superposition of nonlocal correlations that can only be detected by observing both subsystems and, via an ensemble of trials that individually record corresponding outcomes at both subsystems, determining the state of the correlations between them.

Although the global state implies the local description, the local description cannot hint at the global correlations because any such hint would violate Einstein causality.

Thus, when an electron shows up in your lab, neither an examination of the electron nor an examination of an ensemble of identically-created electrons can give you the least hint of whether or how this electron is entangled with other quanta elsewhere in the universe.

This clarification of entanglement resolves the problem of definite outcomes, aka Schrodinger's cat.

An ideal measurement of a superposed microscopic system A by a macroscopic detector B establishes the measurement state (5) at 100% positive correlation.

This state is equivalent to the logical conjunction "A is in local state $|\psi_1\rangle$ if and only if B is in local state $|1\rangle$, AND A is in local state $|\psi_2\rangle$ if and only if B is in local state $|2\rangle$," where AND indicates the superposition.

This conjunction is precisely what we want following a measurement.

Schrodinger's cat is not in the least paradoxical.

Still, this analysis **does not** entirely resolve the quantum measurement problem.

It resolves the problem of definite outcomes associated with the measurement state (5), but this state continues to obey Schrodinger's equation and is hence reversible.

In fact, the entangled state between a quantum and its which-path detector can actually be reversed in the Stern-Gerlach experiment.

In my view, a quantum measurement must result in a macroscopic indication such as a recorded mark, and a mark is **irreversible**.

The above analysis shows the entangled state (5)

describes a mixture of definite, not superposed, outcomes of measurements,

but these outcomes remain indeterminate and the global state remains reversible.

The irreversibility problem is the question of

how this nonlocal superposition of correlations then further "collapses" irreversibly to just one of its possible outcomes, a "collapse" that occurs in the RTO experiment only when one photon impacts a detector.

Present analysis does not seem to resolve this problem - **I think Gambler's Ruin might do it**!

In the case of the RTO experiment, however,

it seems fairly clear that the non-local superposition described by Eq. (5)

must irreversibly decohere when either of its subsystems A or B interacts with a detector.

The RTO experiment furnishes a particularly good setting for this question,

because the two photons remain in the reversible entangled state (5)

throughout their flights from the source to detectors,

and thus the two key questions of the measurement problem

(the problem of definite outcomes and the problem of irreversibility)

can be analyzed individually.

Let us now **attempt** to resolve the seeming existence of an "**irreversiblity**" problem.

The Environment as Monitor

I hope you are by now convinced that quantum physics describes the microscopic world entirely, consistently and with unparalleled experimental accuracy.

But still, there's one slight problem:

Quantum physics seems to fail utterly in describing the world around us!

We never see tables or teapots, not to mention ice cream, as wavy, space-filling fields that are possibly here and possibly there and possibly both here and there.

Tables and teapots don't quantum jump, nor does ice cream.

Classical physics may be inaccurate at the microscopic level, but it does seem to explain how such ordinary objects move in response to forces.

If quantum physics describes the microscopic world correctly, and if the macroscopic world is made of microscopic objects, then the quantum principles should lead ultimately to tables and teapots.

Let us now explain, at least in part, **one way that might work**.

As we have seen, quantum physics began with Max Planck's hypothesis that eventually led to the quantum.

The crucial quantum principle is the universe is made of these highly unified, extended **bundles** of field energy.

To the extent that any given phenomenon depends on the spatially extended field nature of quanta, the phenomenon should be considered as "quantum".

To the extent that quanta can be represented by "pointlike" objects, the phenomenon can usually be considered "classical".

But the existence of a classical-quantum boundary is a useful metaphor, not a law of nature.

Most physicists, me included, regard the world as fully quantum(as you will see later).

The clearest expression of the extended field nature of quanta is the principle of superposition.

Because electrons, photons, atoms, and so on are simply disturbances in fields, these objects superpose just as disturbances in the surface of a tub of water superpose.

Several states can all be present at the same time.

We will describe how the process called **decoherence** converts these superpositions into classical-like mixtures later.

The quantum measurement problem must be part and parcel of any discussion of how quantum physics explains our normal world.

After all, quantum measurements - quantum phenomena that cause macroscopic changes - are the bridge from the micro- to the macroworld.

Schrodinger's cat is a dramatic example.

The microscopic decay of a radioactive atom triggers a device that can kill a macroscopic cat. Schrodinger launched the measurement problem when he noted that the quantum rules seem to imply something we never see: a macroscopic superposition - namely, a cat that is both alive and dead.

Earlier, we suggested a resolution of this problem of definite outcomes. **But there's more than that to the measurement problem.**

Macroscopic processes are **irreversible**. The moving finger moves on. "Things run down". The second law of thermodynamics demands it.

So quantum measurements must be irreversible, even though there appears to be no trace of irreversibility in the microscopic world. As we'll see, decoherence **seems** to solve this mystery.

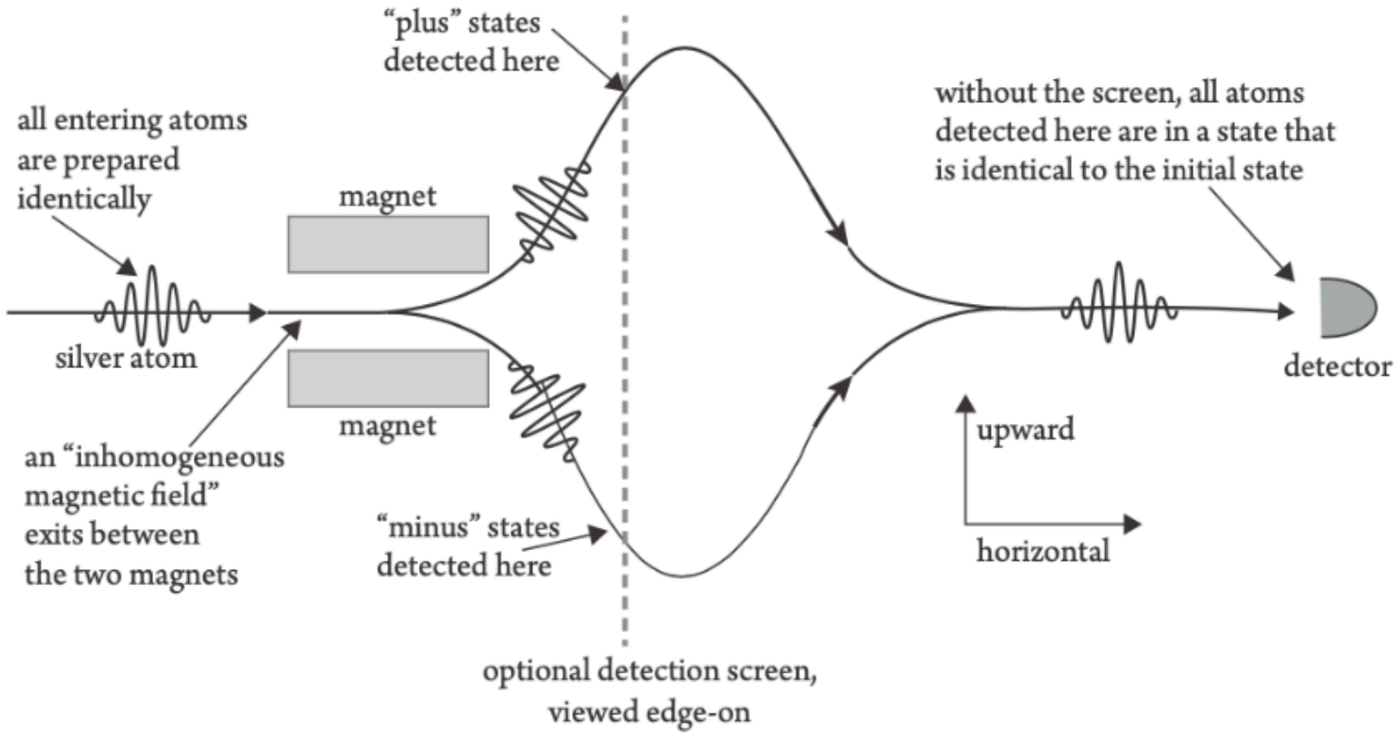
The Problem of Irreversibility

Every measurement involves a macroscopic change of some sort, and such changes must obey the second law of thermodynamics, so entropy increases. Furthermore, we can observe this entropy increase as a permanent (i.e., irreversible) mark made by the measurement.

But when a quantum is measured, the quantum and its detector obey the time evolution equation as they **entangle** to form the measurement state, so it seems that entropy doesn't increase. How can we resolve this conundrum?

A variation on the Stern-Gerlach experiment demonstrates the connection between quantum measurements and the second law. I'll focus only on the features of this experiment relevant to the irreversibility of measurements. Here's the experiment.

A horizontal stream of silver atoms passes between a pair of magnets (see figure).



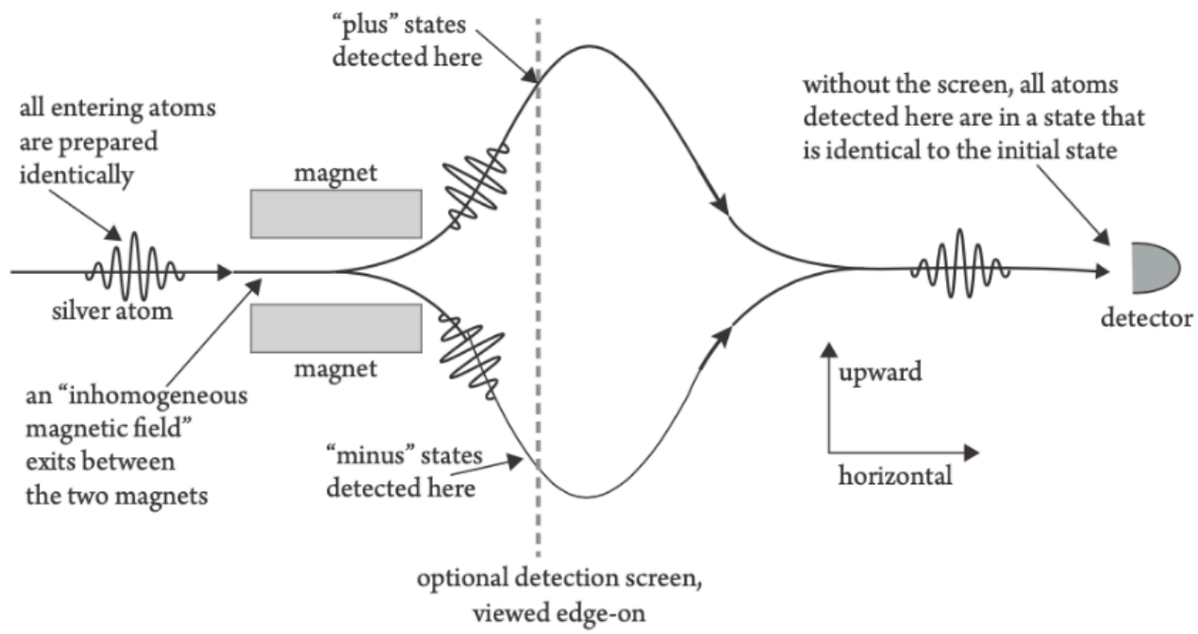
The entering atoms, at the left, have all been prepared previously in identical states that I'll call the "zero state".

There’s no need for us to concern ourselves with the precise nature of this initial state.

The magnets are shaped to create a so-called "inhomogeneous magnetic field" in the space between the magnets, as shown in the figure.

This field separates the stream in such a way that if a detection screen (dashed line) is placed "downstream" from the magnets as shown, atoms make **visible impacts** on the screen at two different spots, one above and the other below the original direction.

Individual atoms impact randomly at one or the other spot with a 50% probability - a striking example of quantum indeterminacy - because the atoms were all prepared identically.



Examination of the two beams shows that atoms striking the upper spot are no longer in the zero state but are in a different state called the "plus state", whereas the atoms striking the lower spot are in a new state called the "minus state".

So the magnet-plus-screen combination acts as a detector to determine which atoms are in the plus state and which are in the minus state. This is entirely analogous to a double-slit experiment with a which-slit detector putting quanta into the "slit 1 state" or the "slit 2 state" before the quanta strike a viewing screen.

There’s more, however.

If one removes the detection screen and instead installs some appropriately chosen magnets (not shown in figure) at certain points along both paths, one can bend each stream back onto its original horizontal path, as shown. **Removing** screen —> **No** measurements made now.

On studying the atoms in the converged stream, a perhaps surprising result emerges; none of these atoms are in the plus state and none are in the minus state. Instead, every one is in the zero state from which it started!

We created a **"do-nothing" box**!

So this experiment, without the detection screen(without the measurement), **is** reversible.

But with the screen, the experiment is obviously **irreversible** because the atoms make an **irreversible** mark when they strike the screen.

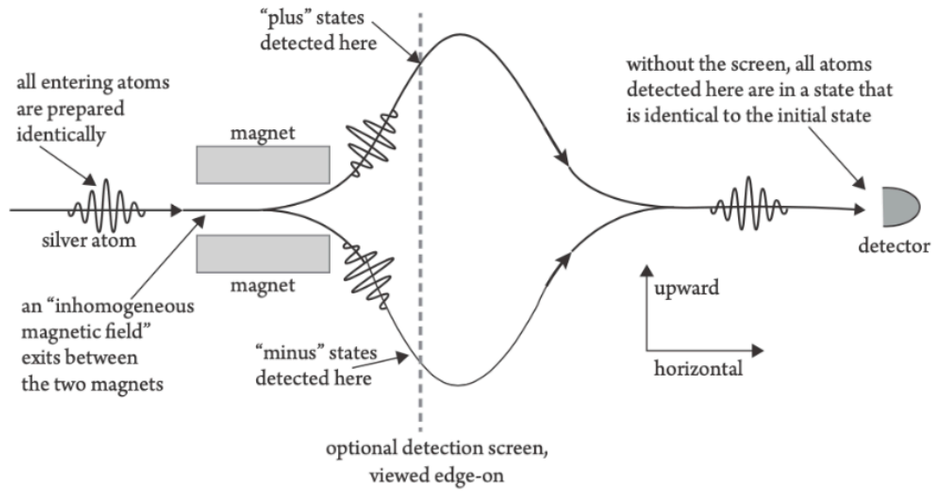
It follows that any irreversibility and corresponding entropy increase in this experiment are **entirely due** to the impact on the screen, because we have seen that, without the screen, the experiment is reversible.

Thus the macroscopic detection makes all the difference, and it is here that we must search for all the irreversible effects of measurement.

Specifically, mere entanglement, such as occurs when a which-slit detector operates in the double-slit experiment, is **not responsible** for irreversibility.

The detection at the screen works as follows:

The detection screen interacts with each atom to form the entangled measurement state: atom is in the plus state, flash appears at the upper spot entangled (superposed) with atom is in the minus state, flash appears at the lower spot



This is the controversial Schrodinger's cat state that we analyzed earlier (the radioactive nucleus is now the silver atom, and the cat is now the screen), where we argued that this state is only a superposition of correlations.

This experiment demonstrates that the macroscopic recording of a measurement is no small detail.

No measurement is complete until it makes a mark on the macroscopic world, and making such a mark must, because of the second law, involve an entropy increase, implying irreversibility.

As John Wheeler repeatedly stressed,

> No elementary quantum phenomenon is a phenomenon until it is brought to a close by an irreversible act of **amplification(so we can see it!!)**

If the measurement is recorded by the audible click of a detector, for example, the irreversible mark is a sound wave spreading out in all directions into the air around the detector.

This sound wave warms the air a little and eventually disperses (vanishes, **for all practical purposes**) into a large volume of air. There is no way nature is going to spontaneously gather up every last bit of this wave's dispersed energy (while necessarily cooling the air), reverse the entire dispersal process, and use this energy to restore the detector and the air to their states before the click.

In fact, the reversed process is prohibited by the second law because it would reduce the total entropy of the universe!

We conclude that microscopic quantum processes, including entanglement, remain reversible as long as they remain microscopic, and that the irreversibility of measurements must be rooted in the macroscopic recording process.

We now look at a macroscopic recording process that goes on all over the universe all the time.

How Environmental Decoherence Collapses Superpositions

A pebble lies on a sunny beach, immersed in an environment that includes atmospheric molecules, photons from the sun, cosmic rays from stars, and even photons from the Big Bang.

During every second, many such quanta interact with the pebble, reflecting or otherwise scattering off the pebble in every possible direction.

The scattered photons must carry away data about the pebble's orientation, structure, and color, because otherwise the pebble could not be seen.

Such natural "measurement" processes occur all the time, regardless of the presence or absence of humans to consciously observe the gathered data.

To study the quantum features of such a natural measurement, instead of a pebble let's consider an atmospheric atom that is in a highly nonclassical state of being superposed at two or more macroscopically separated locations.

We know that such a superposition can be created in the laboratory, for instance by passing the atom through a double-slit apparatus, but it could also occur naturally, for example if the atom passes through an opening that is sufficiently narrow to cause the atom to diffract widely.

What happens when such an atom interacts with quanta in the surrounding environment?

The superposed atom interacts with, say, an environmental photon in a manner analogous to the way a superposed electron coming through double slits interacts with a which-path detector.

If the interaction is significantly different at the atom's two superposed locations, a photon can "measure" the atom by entangling with it in a measurement-like state.

Such an entangling interaction is just like a which-path measurement as discussed earlier!

The scattered photon carries away which-path(location) data about the atom, just as a which-path detector carries away data about an electron coming through the slits.

As discussed earlier(via reduced density operator), this measurement collapses or decoheres (removes, via a series of small environmental interactions, the interference pattern from) the superposed atom, converting its state from a superposition of being at both locations to a local mixture of being either at the first location or at the second location.

In the natural environment, a single scattered photon is not likely to entirely measure (entirely decohere) the atom, because environmental quanta are not specifically organized, the way a laboratory detector is organized, to measure superpositions.

A which-path detector in a laboratory is carefully constructed to respond in detectably different ways at different locations of the measured quantum, quite unlike environmental quanta, which interact randomly in all kinds of ways.

Although a good which-path detector requires only one interaction to reliably distinguish between - reliably decohere - the superposed states of the detected quantum, a large number of environmental quanta must scatter from a typical superposed atom to completely decohere it and turn it into a local mixture.

Careful analysis shows that decoherence of typical naturally occurring superpositions requires many environmental interactions.

So decoherence by the environment **involves** a series of partial measurements and partial collapses, **each** of them instantaneous, nonlocal, and similar to the single-step measurement that occurs at the slits in a double-slit experiment with a which-path detector.

It's through this environmental decoherence process that everyday objects such as pebbles and this slide lose their extended quantum field nature and behave classically, with no obvious trace of superposition, interference, or nonlocality.

Small objects, of atomic dimensions, are less susceptible to environmental decoherence simply because fewer environmental quanta scatter from them as a result of their smaller size, and because each scattering event can cause only a tiny amount of decoherence because the entanglement involves nearly indistinguishable locations.

Thus, a superposed photon from the Big Bang might travel the universe for 13.8 billion years without decohering, while a grain of sand on Earth, should it happen to show any signs of superposition, is decohered environmentally and essentially instantly because of the myriad environmental interactions it experiences.

Widely extended superpositions of small objects, such as a fine dust grain superposed in two locations separated by a millimeter, are also decohered nearly instantly because the branches of such a superposition are so distinct that a single photon reflecting from one branch but not the other can turn the entire superposition into a mixture.

In a similar way, a single detector at only one of two parallel slits is sufficient to turn a quantum that's initially in a superposition of coming through both slits into a mixture that comes through either one or the other slit.

For mesoscopic objects such as dust grains, environmental decoherence turns superpositions quickly into mixtures.

How quick?

This has been calculated theoretically and measured in a few experiments.

A fine dust grain is some 10^{-5} m across.

If it's in a superposition of being in two places with its two superposed branches separated by this same distance so that the two branches are right next to each other, as though the grain had come through two closely adjacent slits, it would be decohered entirely by normal air on Earth in only 10^{-31} second.

Even the best laboratory vacuum (which still contains plenty of air molecules) would decohere it within 10^{-14} second - a hundredth of a trillionth of a second.

If this grain were in deep outer space, cosmic background radiation from the Big Bang would decohere it in 1 second.

For a smaller object, such as a large molecule with a diameter of 10^{-8} m, these decoherence times are longer (still assuming it's in a superposition with two branches that are separated by the molecule's diameter): 10^{-19} second if the molecule is in normal air, 0.01 second if it's in the best laboratory vacuum, but much longer than the age of the universe if it's in deep space.

The message is that sufficiently small superpositions can survive awhile, but meso- or macroscopic superpositions are fragile and are decohered quickly by tiny environmental interactions.

This is why the "quantum world" is usually identified with the microworld.

When typical quantum features become meso- or macroscopic, they generally vanish quickly.

So the quantum universe appears classical at the macroscopic level because the enveloping quantum environment "monitors" every object constantly, and macroscopic objects are especially susceptible to this decoherence process.

Nature is full of which-branch detectors! Note that humans aren't required in any of this - no physicists, no laboratories needed.

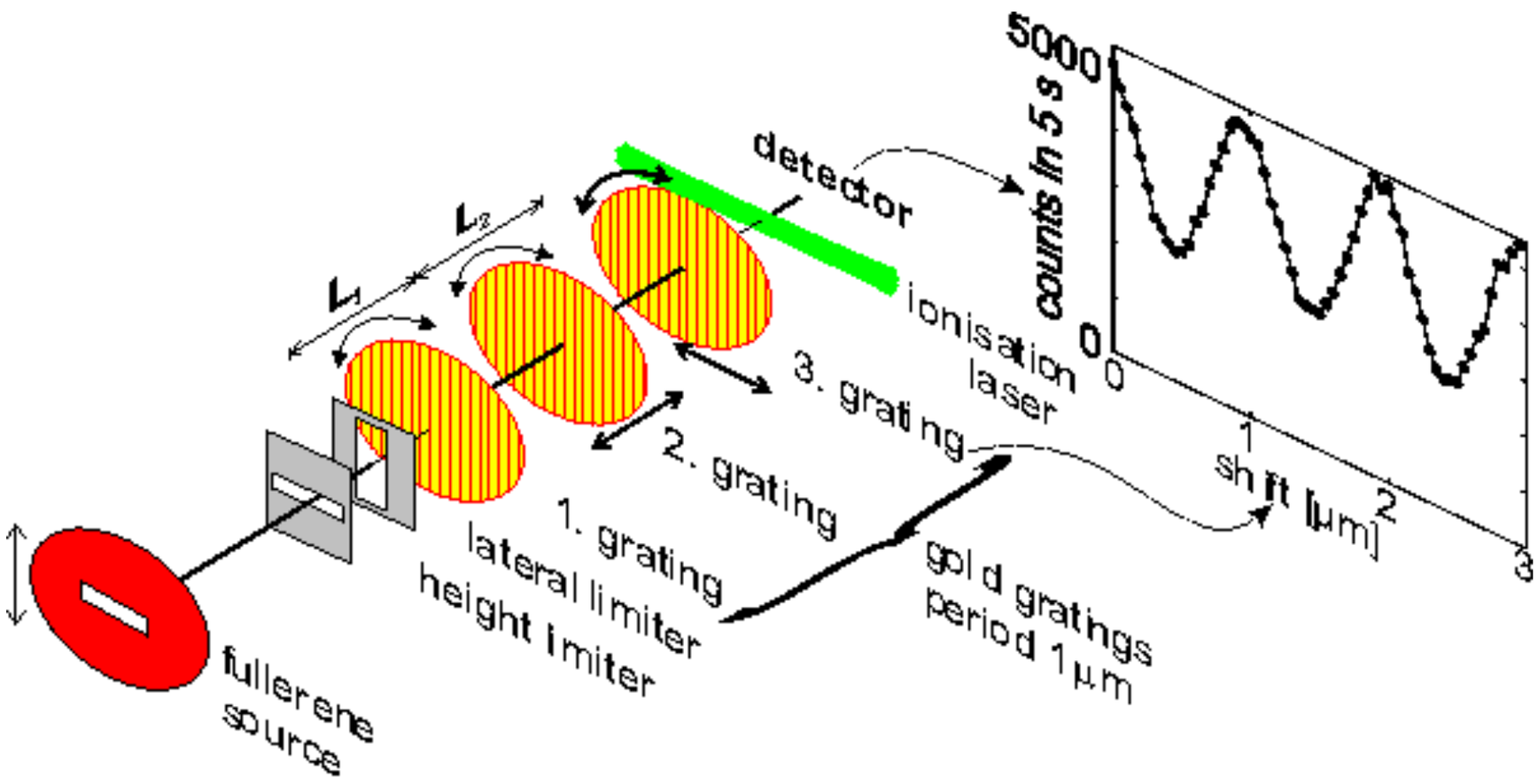
Nature has been collapsing superpositions, and quanta have been losing their quantum field nature, at least since the Big Bang.

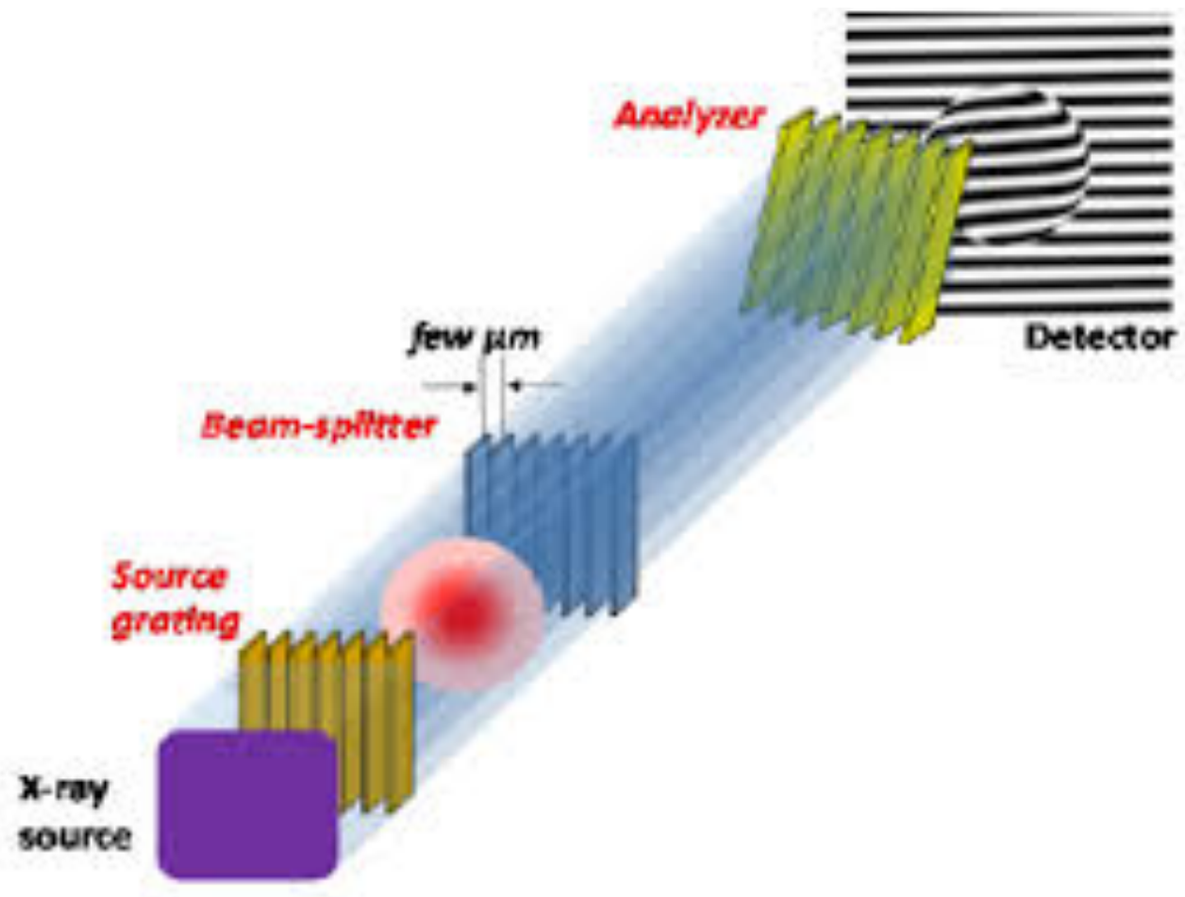
The role of the environment as an ever-present which-path monitor that turns mesoscopic and macroscopic superpositions into mixtures was first clearly recognized by Zurek during the early 1980s. The work of Zurek, his colleagues, and others has, in large part, explained how the quantum world leads(via reduced density operators) to the classical world of our experience.

A wide variety of experiments have demonstrated environmental decoherence and convinced physicists that decoherence really is the mechanism that converts the quantum world into tables and teapots.

One beautiful example is an experiment by the University of Vienna group under Zeilinger and Arndt. They use the Talbot-Lau interferometer technique that demonstrated, in 2002, interference in large molecules such as C70 and certain biological molecules. The interference showed each of these large molecules to be in a superposition of following more than one path on its journey through the interferometer. Using the same technique in 2004, this group was able to demonstrate convincingly the "environmental" decoherence of superpositions of C70 molecules.

I put environmental in quotation marks because in this experiment the environment came from within the molecules themselves rather than from an external environment. Individual C70 molecules passed through the interferometer, as in the 2002 experiment.


detector
ionisation laser
3. grating
2. grating
1. grating
lateral limiter
height limiter
fullerene source
gold gratings
period 1 µm
5000
counts in 5 s
0
0
1
2
3
shift [µm]


Analyzer
Detector
few µm
Beam-splitter
Source grating
X-ray source

But there was a new twist; the experimenters heated the molecules just before sending them through the interferometer.

They expected that, with sufficient heating, the molecules would themselves emit thermal radiation in the form of visible and infrared "thermal photons", just as an electric hot plate emits thermal radiation (you can feel its warmth at a distance, and it might glow red) when heated.

The radiation comes from the random thermal motion of the many atoms and other quanta within the molecule.

According to decoherence theory, each of these radiated photons should act as a partial which-path detector, collapsing the molecule's superposition state partially by carrying a certain amount of which-path data from the molecule into its surroundings. If enough such data are transferred, this which-path measurement should decohere the molecule, causing the interference pattern to dim or vanish - the signature of a superposition evolving into a mixture.

Their results demonstrated decoherence in action.

With the molecules only slightly heated, the emitted photons had low energies and thus low frequencies and long wavelengths - too long to distinguish between the possible paths, which for such a massive molecule are separated by extremely small distances.

But when the molecules were heated to a few thousand kelvins, the emitted photons' wavelengths became short enough to distinguish between the possible paths, so the which-path data transmitted to the surroundings was sufficient to partially decohere the superposed molecules, causing the interference pattern to partially vanish.

At sufficiently high temperatures, theory and experiment showed that a mere two or three emitted photons sufficed to decohere each molecule.

There was quantitative agreement between the theory predictions and observations.

The interference pattern began decohering at just the predicted temperature, and the degree of visibility of the remaining interference was just as predicted.

The experiment revealed the step-by-step action of decoherence as data leaked, photon by photon, into the environment.

It also demonstrated the extreme sensitivity of superpositions to decoherence.

Just a few high-temperature photons turn a massive superposed molecule into an incoherent mixture.

Superposed quanta are very delicate.

Decoherence and the Measurement Problem

Decoherence provides a solution of the irreversibility problem for **natural** environmental measurements.

Decoherence was first introduced during the early 1980s to explain how our apparently classical surroundings arise from measurement-like interactions with the environment.

The local state argument, presented earlier, shows that such measurements turn superpositions into entangled nonlocal measurement states with definite outcomes, resolving the problem of definite outcomes for environmental measurements.

But just as for laboratory measurements, we must ask if this also resolves the irreversibility problem associated with **natural** measurements.The answer is yes.

Here’s why.

As we've seen, when a small superposed dust grain is decohered by the surrounding environment, the object undergoes a series of small entanglement-caused partial collapses.

These environmental measurements are "recorded" by the many environmental photons and air molecules that interact with the grain and then disperse widely.

Data about the superposition state of the grain before it decohered are now scattered randomly far and wide.

Just as one cannot unscramble an egg, one cannot reversibly gather these pieces back together and reconstruct the global state of the air plus the superposed grain.

This argument is especially compelling in the case of environmental decoherence, because of the environment's enormous size.

This is an obvious example of the second law in action; **for all practical purposes**, the process is irreversible and entropy has increased.

Because there is famous opposition to introducing **for-all-practical-purposes** arguments into physics, it needs to be noted that the second law itself is inherently a **for-all-practical-purposes** principle.

We know that a box full of gas could evolve spontaneously into separate regions of hot gas and cold gas simply by chance, with no external assistance.

The chances of this are ridiculously small, but they are not zero, and if we consider boxes containing smaller and smaller amounts of gas, these odds increase.

Such **for-all-practical-purposes** arguments that trace back to the second law are entirely in keeping with the principles of physics.

So the local state resolution plus decoherence combine to resolve the measurement problem entirely for the case of natural environmental measurements.

The environment measures a superposed object (molecule, dust grain, and so on) when myriad environmental quanta convert the superposition into a measurement state that now takes the form of a nonlocal entanglement between the object and a very dispersed state of the many environmental quanta.

The local state argument implies this measurement state represents a mixture of definite properties, whereas the highly dispersed nature of the environmental quanta guarantees the process is irreversible.

There has, for years, been a question about what is the exact relationship between decoherence and the solution of the measurement problem.

Some accounts appear to imply that decoherence alone resolves the measurement problem, but this is not true.

The connection of decoherence to the measurement problem is that it resolves the irreversibility problem in the case of natural environmental measurements, and it shows how environmental interactions transform a superposed quantum into the measurement state, but it does not solve the problem of definite outcomes associated with this measurement state.

The problem of definite outcomes is, however, resolved by the local state analysis, as explained earlier.

So the local state solution combines with environmental decoherence to solve the measurement problem for the case of environmental measurements.

What about the case of laboratory measurements?

Here, the preceding discussion resolved the definite outcomes problem, but not the irreversibility problem.

Irreversibility poses slightly different problems in the two cases, because lab measurements are sufficiently controlled that the natural environment has little effect(by design).

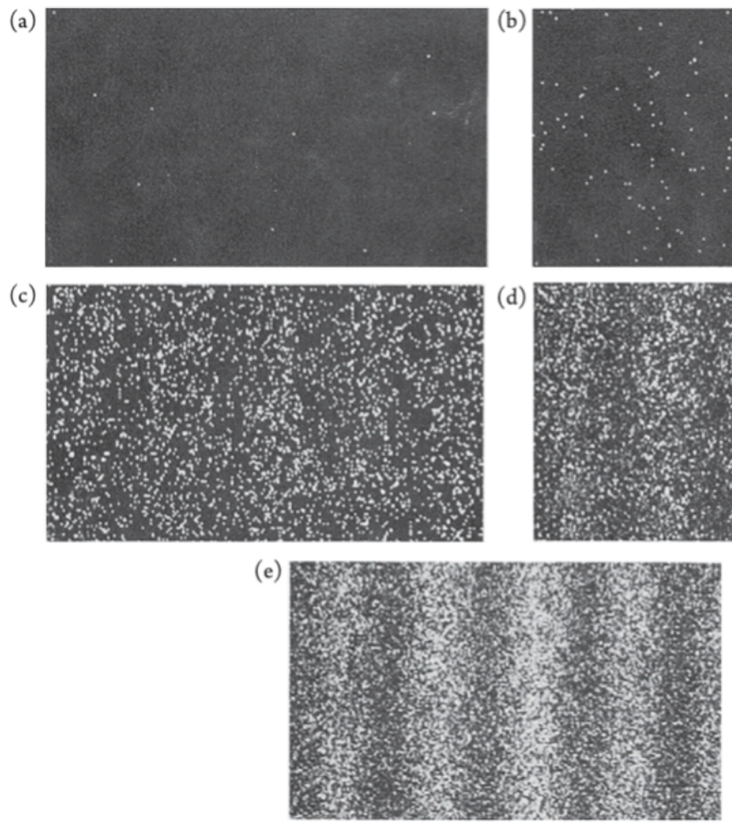
In fact, much of an experimentalist's efforts go precisely into ensuring that random environmental interactions have no significant effect on an experiment's outcome.

Our discussion of the Stern-Gerlach experiment suggested that the answer lies in the irreversible nature of the macroscopic detection process.

This suggestion resembles the resolution of the irreversibility problem for environmental measurements:

Environmental measurements are "recorded" by innumerable environmental quanta whereas lab measurements are recorded by detection screens, electronic clicks, or, perhaps, cats.

Let's focus on a single impact made on a detection screen by one electron in a double-slit experiment - one of the small spots in the figure below:



(a–e) Buildup of an interference pattern in the double-slit experiment using electrons.

As in a regular interference experiment, an interference pattern builds up from individual impacts over longer-time exposures.

How was it recorded?

This mark was initiated by a single electron, but to observe it macroscopically the original impact had to be detected and then **amplified** sufficiently for humans to see it.

This is the purpose of every laboratory detector of microscopic events, and it's why an **"irreversible act of amplification"** (as Wheeler puts it) is essential to quantum measurements.

In the 1989 experiment that produced the patterns of dots on a screen, each electron was emitted in a single coherent state, the same state for each electron.

EM fields accelerated these electrons to high speeds, about 180,000 kilometers per second, or 60% of light speed - much faster than the electron in a normal hydrogen atom, which orbits at an average speed of "only" 2000 kilometers per second.

High energies were needed to make a detectable impact.

Each electron went through a pair of parallel slits, then spread out into an interference state that interacted indeterminately with the screen.

The interaction entangled the electron with the screen, decohering the electron and collapsing it instantaneously from its earlier superposition state over the entire screen to a locally observed mixture(a flash).

In other words, the electron collapsed randomly into one or the other of many small atom-size regions in the screen.

For the fluorescent screen used in the experiment, this interaction created some 500 photons that marked and amplified the location of the impact - already an irreversible process because the photons are emitted randomly in all directions.

In a process called the photoelectric effect, each photon then struck a metal plate and dislodged an electron from it, converting each of the 500 photons into a low-energy electron.

These 500 low-energy electrons were then focused into a point image that could be displayed on a TV monitor in much the same way that old (before flat screens) TV tubes operated.

Every step of this process - creation of 500 photons, conversion to photoelectrons, and amplification to create the final display - creates entropy and is irreversible thermodynamically.

It’s the second law of thermodynamics in action.

This illustrates the general case:

Laboratory detection and amplification of a quantum event necessarily involves irreversibility and entropy production.

The laboratory detector, a macroscopic device made of many microscopic quanta, plays the decohering role that the environment plays in natural measurements.

To summarize:

The quantum measurement problem, in both its laboratory and environmental senses, seems to be resolved entirely by the local state solution of the problem of definite outcomes and by the irreversibility of the decoherence process.

For more than a century, much has been made of the odd and supposedly paradoxical nature of the quantum.

This presumed quantum spookiness has led to an excess of attempted fixes and interpretations.

Many experts have even declared the theory to be not a description of reality at all, but only a mathematical recipe that helps humans predict the results of experiments.

As Niels Bohr put it, **“There is no quantum world. There is only an abstract quantum description.”**

According to this hypothesis, quantum theory doesn’t describe anything real at all, so there’s no cause for concern about collapse of the quantum state and other odd quantum behaviors.

We now collect everything we have proposed together, eliminate all that is unneeded and finally present a straightforward picture of the solution to the quantum measurement process

We will see that the final understanding arises because we looked at so many different approaches, went down so many dead-ends and along the way uncovered many new features of the quantum world.

So, now we will present only what we need with a more detailed analysis and a deeper and, what I believe, is a correct interpretation.

1 - INTRODUCTION

Physicists agree that Schrodinger's equation describes the unitary evolution of non-relativistic quantum states between measurements, but there is no agreement on how states change during measurements.

In fact, an apparent problem arises when one applies standard quantum theory (minus the collapse postulate) to measurements.

John von Neumann, in 1932, analyzed the problem and we follow his argument here.

> "Measurement" is the **experimental determination and macroscopic recording** of the value of a physical observable associated with a quantum system.
>
> von Neumann showed that, **unless** the system happens to be in an eigenstate of the measured observable, measurement leads to a **"measurement state" (MS)** whose mathematical representation is an **entangled state** that **seems to predict** the detector to be in a **macroscopic superposition** of exhibiting all of the possible outcomes, a paradox known as the **"problem of definite outcomes."**

We will now show, based on standard quantum theory without a collapse postulate, that this is a **pseudo-problem** and that, far from predicting superposed outcomes, von Neumann's MS **predicts an instantaneous collapse** to a single eigenstate.

Specifically, we will demonstrate, with no assumptions other than standard quantum physics (minus the collapse postulate)

(i) The MS has been **misinterpreted** (for 90 years) and **does not** in fact predict paradoxical multiple macroscopic *outcomes;* it instead correctly predicts non-paradoxical multiple ***statistical correlations between*** system and detector outcomes.

(ii) Exactly **one** outcome actually occurs.

(iii) The entanglement says that, when one outcome occurs, the **other outcomes** simultaneously and nonlocally remain "dark" (i.e. **do not occur**).

This **resolves** an objection to quantum physics first raised by Einstein in 1927.

That is, we will show von Neumann's enigmatic MS **to be in fact** the collapsed state expected upon measurement.

The collapse is derived (i.e. demonstrated) with **no assumptions** beyond the other standard principles of quantum physics(i.e., without a collapse postulate!).

We will show the problem of outcomes arises from a **mistaken understanding** of entangled superpositions, **not only** in measurements **but also** in purely microscopic processes.

The MS is obviously a **superposition** of subsystem product states.

But the MS is poorly understood because no previous work has analyzed its complete **phase dependence**. Remember our earlier discussion of state vectors for boxes of electrons!

Do the states of individual subsystems vary with phase, as they do in simple superpositions?

If not, then what does vary with phase, i.e., precisely which entities occupy indefinite states when two or more subsystems are entangled?

Such questions show that we do not fully understand the MS's phase dependence, i.e., we don't fully understand the MS.

We investigate these questions by studying earlier Bell-test quantum-optics experiments that measure momentum-entangled two-photon states.

These experiments study a **purely microscopic** entangled superposition (mathematically **identical** with the MS) **across all phases**.

The results show that entangled superpositions **differ sharply** from simple superpositions, i.e., the implication of the **"plus"** sign differs surprisingly.

The MS **does not** represent multiple detector readings, but instead represents multiple ***statistical correlations between*** system states and detector readings.

These correlations are **not paradoxical**: a macroscopic detector that simultaneously exhibited two states **would be paradoxical**, but a detector that simultaneously participates in two correlations is **not paradoxical**.

Furthermore, the MS **directly implies** that **exactly one** of these correlations is **realized** as the measurement **outcome**.

This **resolves** the so-called "Schrodinger's cat paradox," which in turn **resolves** the quantum measurement problem.

Our discussion follows from quantum theory and from experimental evidence(already discussed) that an instantaneous nonlocal collapse takes place, **resulting in one outcome occurring while the other outcomes simultaneously do not occur**.

Thus, while the present discussion **does not *postulate* collapse**, it ***derives an instantaneous collapse*** as a **consequence** of **entanglement**.

Many papers, and our earlier discussions, express the effect of measurement as a **trace** on the composite-system density operator representing the MS.

This trace operation, which **predicts definite outcomes** at the subsystems, seems to yield just what we want, namely that measurement **transforms** the subsystem states into **mixtures** over the possible eigenstates.

However, **several** objections are commonly raised against this proposal, so we now go into more detail to understand what is the **correct** interpretation that overcomes objections.

This discussion, in contrast to earlier discussions, argues that the MS **directly** represents the collapsed state of the composite system, that the collapse is a **consequence** of entanglement, and that the entanglement is **required** in order to ensure a **simultaneous** (hence instantaneous) collapse over the separated branches of the experiment.

We will **present** von Neumann's derivation of the MS, **pose** the measurement problem, **review** eight presumed measurement problem insolubility proofs, and **explain** why their conclusion cannot be correct.

We then present a **crucial** clue — In 1927, **Einstein** noted that quantum theory **implies** that measurements entail **instantaneous** collapse and suggested this would **violate** Special Relativity.

However, today (unlike 1927) we **know** that **instantaneous** nonlocal changes of **correlations** violating Bell's inequalities really **occur**, that they **do not violate** special relativity, and that they occur when disparate systems are **entangled**. They have been observed experimentally!

Thus, from today's perspective, Einstein's **argument** implies that **nonlocal correlations** are **required** during measurements and also saying that **entanglement** is also **required**.

Is von Neumann's MS in fact precisely what we **want**?

The answer can only come from **fully understanding** the MS, particularly its **full phase dependence**. To this end, we will review two 1990 quantum optics experiments exploring an entangled two-photon microscopic state that is **mathematically identical** to the MS, that we discussed earlier.

The **correct** understanding of the microscopic version of the MS will then be worked out.

We **find** that both entangled sub-systems are in **definite** (i.e. non-superposed) states, but that the ***degree of correlation between these states is in an indefinite state***.

We will then apply this new **insight** to the MS.

We will find the MS is **not** a macroscopic superposition of different detector *states*.

It instead **represents** *different correlations between* detector states and system states.

This is **not paradoxical**.

Furthermore, quantum theory directly implies that precisely **one outcome is realized**.

This **resolves** the Schrodinger's cat paradox.

We will then conclude the analysis by studying a particularly **simple measurement example** that also typifies the **essentials** of quantum measurements.

Finally we will summarize the results and consider **why it has taken so long to overcome this simple misunderstanding of von Neumann's MS**.

2 - ENTANGLEMENT AND MEASUREMENT

The superposition postulate entails that, if $|A1\rangle$ and $|A2\rangle$ are Hilbert space vectors ("kets") representing possible states of a quantum system *A*, then all normalized linear superpositions of $|A1\rangle$ and $|A2\rangle$ also represent possible states of *A*. For example,

$$|\Psi_A\rangle = \frac{1}{\sqrt{2}}(|A1\rangle + |A2\rangle) \tag{1}$$

You should make a copy of all the numbered equations for reference as I lecture

($|A1\rangle$ and $|A2\rangle$ are orthonormal) represents a possible state of *A*.

The superposition postulate is prerequisite to the Hilbert-space representation of quantum states, and the basis for conceptualizing quantum states as physically real waves in a quantum field that fills the universe.

$|\Psi_A\rangle$ represents a situation in which *A* is represented neither by $|A1\rangle$ nor by $|A2\rangle$ but incorporates aspects of both, including "overlap" effects such as interference.

As **Dirac** put it, *A* goes "partly into each of the two components" and "then interferes only with itself."

If quantum system *A* interacts with another quantum system *B*, it frequently happens that the situation of *A* and *B* are then represented by an entangled superposition state vector such as

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}}(|A1\rangle\,|B1\rangle + |A2\rangle\,|B2\rangle) \tag{2}$$

where $|Ai\rangle$ and $|Bi\rangle$ (*i = 1, 2*) are orthonormal kets representing the "subsystems" *A* and *B*, respectively.

Although the physical interpretation of simple superpositions such as $|\Psi_A\rangle$ is clear, the physical interpretation of entangled superpositions such as $|\Psi_{AB}\rangle$ is not comparably clear.

$|\Psi_{AB}\rangle$ is a superposition of two products $|A_i\rangle\,|B_i\rangle$ (i =*1, 2*). $|A_1\rangle\,|B_1\rangle$ is commonly interpreted to represent a state of the composite system *AB* in which *A* has the properties associated with $|A1\rangle$ and *B* has the properties associated with $|B1\rangle$, and similarly for $|A2\rangle\,|B2\rangle$.

But if this is the case, then the physical interpretation of the state $|\Psi_{AB}\rangle$ would seem to be that *AB* simultaneously exhibits properties associated with $|A1\rangle$ and $|B1\rangle$ AND properties associated with $|A2\rangle$ and $|B2\rangle$, where "AND" represents the superposition.

In the case of Schrodinger's iconic cat, this would imply that the nucleus is both decayed and undecayed and, more disturbingly, the cat is both alive and dead.

Here we will demonstrate that both quantum experiment and quantum theory show that this is actually *not* the case.

Instead, the state $|\Psi_{AB}\rangle$ entails merely that $|A1\rangle$ and $|B1\rangle$ are *coherently* (in a phase-dependent manner) *correlated* with each other AND $|A2\rangle$ and $|B2\rangle$ are coherently correlated with each other (see later discussion).

This is not paradoxical in any way.

Quantum measurements are important examples of entanglement.

As first discussed by John von Neumann, a "measurement" is the determination of the value of an observable associated with a quantum system *A*.

If *A* happens to be represented by an eigenvector of the measured observable, then a good measurement will detect the associated eigenvalue.

But what if *A* is represented by a superposition of eigenvectors of the measured observable?

It will suffice for this discussion's purpose to assume that *A*'s Hilbert space has only two dimensions, and that *A* is represented by the superposition Equation (1).

The kets $|Ai\rangle$ *(i =1, 2)* define the eigenvectors of the measured observable.

We assume the existence of a **detector** *B* designed to **distinguish** between the $|Ai\rangle$.

For example, $|A1\rangle$ and $|A2\rangle$ could represent the paths of an electron passing through the slits of a double-slit apparatus, and *B* could be an electron detector for the "which-slit" observable whose eigenvectors are the $|Ai\rangle$.

To make a which-slit measurement, *B* must distinguish between the states represented by $|A1\rangle$ and $|A2\rangle$, so *B* must contain a specific quantum detection component with quantum states represented by kets $|Bi\rangle$ such that, if *A* is in the state represented by $|Ai\rangle$, then detection yields the state represented by $|Bi\rangle$*(i = 1 or 2)*.

Assuming a minimally-disturbing measurement that leaves eigenstates unaltered, and letting $|B_{ready}\rangle$ represent the state of *B*'s quantum component prior to measurement, the process

$$|Ai\rangle\,|B_{ready}\rangle \Rightarrow |Ai\rangle\,|Bi\rangle \quad (i = 1, 2) \tag{3}$$

describes a measurement of the which-slit observable when *A*'s state is represented by an eigenstate.

When *A* is in the state represented by $|\Psi_A\rangle$ and *B* measures the which-slit observable, simple linearity of the time evolution implies

$$\frac{1}{\sqrt{2}}(|A1\rangle + |A2\rangle)\,|B_{ready}\rangle \Rightarrow \frac{1}{\sqrt{2}}(|A1\rangle\,|B1\rangle + |A2\rangle\,|B2\rangle) = |\Psi_{AB}\rangle \tag{4}$$

Thus von Neumann's straightforward argument shows the measurement creates the entangled superposition $|\Psi_{AB}\rangle$ of Equation (2), where "*B*" now refers to the quantum detection component of the detector.

But von Neumann's measurement postulate implies that, when the which-slit observable is measured, *A* collapses into one of its eigenstates while *B* collapses into the corresponding detector state.

It is by no means clear that $|\Psi_{AB}\rangle$ (Equation (4)) represents such a measurement outcome.

As many physicists put it, "The problem of what to make of this is called the 'measurement problem'."

This discussion will show that $|\Psi_{AB}\rangle$ does in fact represent the collapsed state and the single definite outcome expected from von Neumann's measurement postulate.

We have used the same notation, $|\Psi_{AB}\rangle$, for the arbitrary entangled state represented by Equation (2) (where *A* and *B* are arbitrary quantum systems) and for the specific case of the entangled state that develops when a detector measures a quantum system, represented by Equation (4) (where *B* is now a detector).

We will refer to this state in the context of Equation (4) as the "measurement state" (MS).

We will also, however, need to refer to the arbitrary entangled state Equation (2), especially later where we analyze an experiment involving two microscopically entangled photons.

Thus the question of how to interpret entangled states looms large in the foundations of quantum physics.

As noted above, the interpretation of general entangled states such as the one represented by Equation (2) is already murky as compared with the interpretation of simple superposition states such as the one represented by Equation (1).

The problem of interpreting the MS is especially important, because **macroscopically distinct** states now lie on each side of the "plus" sign on the right-hand side of Equation (4).

As already discussed, the superposition $|\Psi_A\rangle$ can be interpreted to represent a situation in which *A* incorporates properties represented by both $|A1\rangle$ and $|A2\rangle$.

And a product state such as $|A1\rangle|B1\rangle$ represents a state of the composite system *AB* in which *A* is represented by $|A1\rangle$ and *B* is represented by $|B1\rangle$.

Thus $|\Psi_{AB}\rangle$ appears to describe a detector that simultaneously "points" to two macroscopically different outcomes $|B1\rangle$ and $|B2\rangle$!

The detector seems to display no definite outcome, a conundrum known as the **"problem of outcomes"**

Such a superposition state would be paradoxical.

Schrodinger, who imagined a cat attached to the detector in such a way that $|B1\rangle$ included a live cat and $|B2\rangle$ included a dead cat, described $|\Psi_{AB}\rangle$ as representing a “living and dead cat ...smeared out in equal parts.

As one quantum foundations expert writes,

The crucial difficulty is now that it is not at all obvious how one is to regard the dynamical evolution described by [Equation (4)] as representing measurement in the usual sense.

This is so because [Equation (4)] is ...not sufficient to directly conclude that the measurement has actually been completed.

In fact, while measurement should lead to a specific eigenstate of the measured observable, Equation (4) appears to entail that the system has been **sucked into a vortex of entanglement** and **no longer** has its own quantum state.

On top of that, the entangled state **fails to indicate** any particular measurement outcome.

As noted, it seems paradoxical that quantum measurements lead to a state represented by the MS.

Measurement should lead to a situation in which *A* is represented by one of its eigenvectors $|Ai\rangle$ and *B* is represented by the corresponding $|Bi\rangle$.

Since quantum uncertainty typically implies **unpredictable** measurement outcomes, it is reasonable to conclude that measurement should lead to a state represented by an **ignorance-interpretable mixture** of the products $|A1\rangle|B1\rangle$ and $|A2\rangle|B2\rangle$.

Assuming the initial state is represented by $|\Psi_A\rangle$, such a post-measurement mixture would be represented by the density operator

$$\rho_{mixed} = \frac{1}{2}(|A1\rangle|B1\rangle\langle B1|\langle A1| + |A2\rangle|B2\rangle\langle B2|\langle A2|) \tag{5}$$

This mixture can be interpreted(as we did earlier) as "**the system is represented by a single component** $|Ai\rangle|Bi\rangle$, **but we cannot know whether *i = 1* or *2* until we look at the outcome.**"

Beginning with von Neumann's analysis, at least eight published "**measurement problem insolubility proofs"** have assumed that, in order to obtain definite outcomes, the measurement process should transform the composite system *AB* into a mixture analogous to Equation (5) **(so have we so far!!)**

The initial state of *A* is assumed to be pure and to be represented by a superposition such as Equation (1).

The analysis then investigates whether a suitable composite-system post-measurement mixture can be reached via a unitary process.

To achieve this, the detector must be represented by a mixture initially, because unitary processes **cannot turn a pure state into a mixture**.

Since *B* is macroscopic, such an initial mixture **seems appropriate**.

Thus von Neuman and seven succeeding analysts asked:

Is there an initial mixed-state density operator ρ_{ready} of *B* and a unitary process *U* acting on *AB* such that *U* transforms the initial composite density operator $|\Psi_A\rangle\langle\psi_A| \otimes \rho_{ready}$ into the desired composite mixture?

The eight insolubility proofs, with varying assumptions, say the answer is "no," presumably demonstrating the measurement problem to be insoluble(according to them!).

We will show later, however, that the premise of these insolubility proofs, namely that Equation (5) represents the appropriate post-measurement state, was doomed from the start, precisely because it is *not* entangled and thus cannot have the properties required if quantum theory is to describe the measurement process. **That statement represents the crucial point!!**

To put this another way, there are reasons why the post-measurement state *must* be an entangled state, which implies that it *cannot* be a mixture such as Equation (5), which just represents what we measure!

We will then show that the MS does have the desired properties.

3 - A CRUCIAL CLUE FROM EINSTEIN

At the 1927 Solvay Conference, five years prior to von Neumann's analysis of quantum measurement, Einstein asked the audience to consider an experiment in which electrons pass through a tiny hole in an opaque screen and then impact a large hemispherical detection screen centered at the hole (Fig. 1).

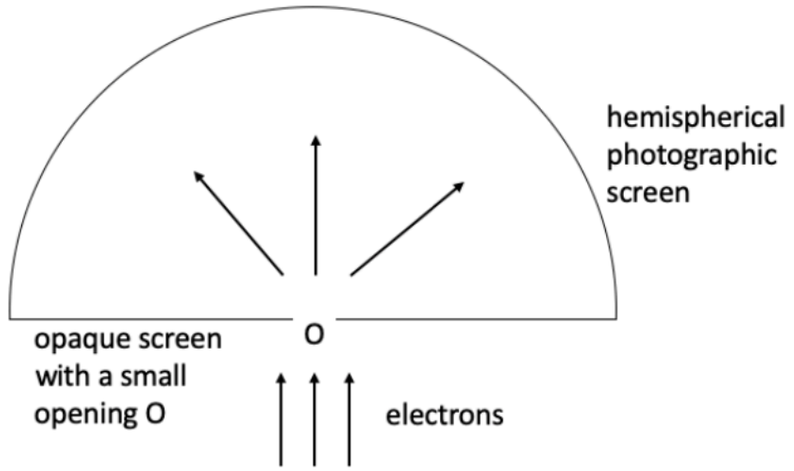


Figure 1. Einstein's thought experiment. Each electron diffracts widely, reaching the entire screen simultaneously, yet only one point shows an impact. How do the other points *instantaneously* remain dark? Does this violate Special Relativity?

According to the Schrodinger equation, each electron's state diffracts widely, spreading and reaching the entire screen simultaneously.

Yet each electron registers at only a **single point**.

How, Einstein asked, do the other points ***instantaneously*** **remain "dark," i.e., not show an impact**?

As Einstein put it in his notes, this "**entirely peculiar mechanism of action-at-a-distance, which prevents the wave continuously distributed in space from producing an effect in *two* places on the screen,"** presents a fundamental problem.

It appears to imply instant signaling, **violating** special relativity.

Einstein's argument shows that, under a **realistic and objective** (**independent of humans**) interpretation of quantum physics, the Schrodinger equation is **at odds** with experimental facts **unless** the electron's state collapses instantaneously and nonlocally upon measurement.

Thus, realistic quantum physics **implies** instantaneously-established nonlocal correlations are **essential** to the *measurement process*.

Indeed, modern experiments have **verified the nonlocal character** of the measurement transition.

Since nonlocality is essential to measurement, the presumed post-measurement mixed state(5) was doomed from the start precisely because it **does *not* exhibit** the required nonlocality.

But entangled superpositions **do exhibit** the required nonlocality.

So from our **modern point of view**, Einstein's argument shows that **entanglement, far from being an unwelcome paradox, is *required* in measurements.**

This is a **crucial clue** and good news for quantum foundations, because von Neumann's predicted MS is **just such an entangled superposition**!

One particular argument, at least, should be mentioned. Its resolution of the measurement problem "diagonalizes the density operator" as we discussed earlier on one of our adventures.

They form the exact density operator $\rho = |\Psi_{AB}\rangle\langle\Psi_{AB}|$, which can be written

$$|\Psi_{AB}\rangle\langle\Psi_{AB}| = \rho_{\text{diagonal}} + \rho_{\text{off-diagonal}} \tag{6}$$

where $\rho_{diagonal} = \rho_{mixed}$ (Equation (5)) and

$$\rho_{off-diagonal} = \frac{(|A1\rangle|B1\rangle\langle B2|\langle A2| + |A2\rangle|B2\rangle\langle B1|\langle A1|)}{2}. \tag{7}$$

Recall that the exact expectation value of any observable F is

$$\langle F\rangle = \text{Tr}(\rho F) = \sum_j \sum_k \rho_{jk} F_{kj} \tag{8}$$

where ρ_{ijk} and F_{ijk} are matrix elements of ρ and F.

This idea argues that off-diagonal terms in Equation (8) can be ignored because they involve matrix elements such as $\langle B1|\langle A1|F|A2\rangle|B2\rangle$ that are non-zero only for a **"fantastic"** observable F because $|B1\rangle$ and $|B2\rangle$ represent the states of **widely separated detectors**.

It says that matrix elements for such **fantastic** observables can, **for all practical purposes**, be neglected so that we can replace ρ by ρ_{mixed}.

But we have seen that this premise is doomed because ρ_{mixed} lacks the required nonlocal properties, so this type of proposal fails (it leaves our original discussions in limbo for the moment also!!).

4 - EXPERIMENTAL STUDIES OF STATES HAVING ENTANGLED SPATIAL PATHS

Earlier we presented the measurement problem and some previous thoughts on the problem.

We now will present a suggested resolution of all difficulties.

We now review interferometry experiments and theory that investigate the *microscopic* entangled superposition $|\Psi_{AB}\rangle$ Equation (2) over its full 0 - to - π range of phases, for a system of two momentum-entangled (i.e. path-entangled) photons.

The results provide a key insight into solving the measurement puzzle.

As preparation, we first study the simple superposition Equation (1).

Consider the interferometer experiment of Figure 2.

On each experimental trial, a photon enters a 50-50 beam splitter *BS1* which transforms the photon's state into the superposition Equation (1) where $|A1\rangle$ and $|A2\rangle$ respectively represent paths 1 and 2.

A series of single-photon trials probes this state using mirrors *M* that bring the two branches together, phase shifters ϕ_1 and ϕ_2 that lengthen the two paths by phases ϕ_1 and ϕ_2, and a second beam splitter *BS2* that mixes the branches together.

Measurement occurs at photon detectors *B1* and *B2*.

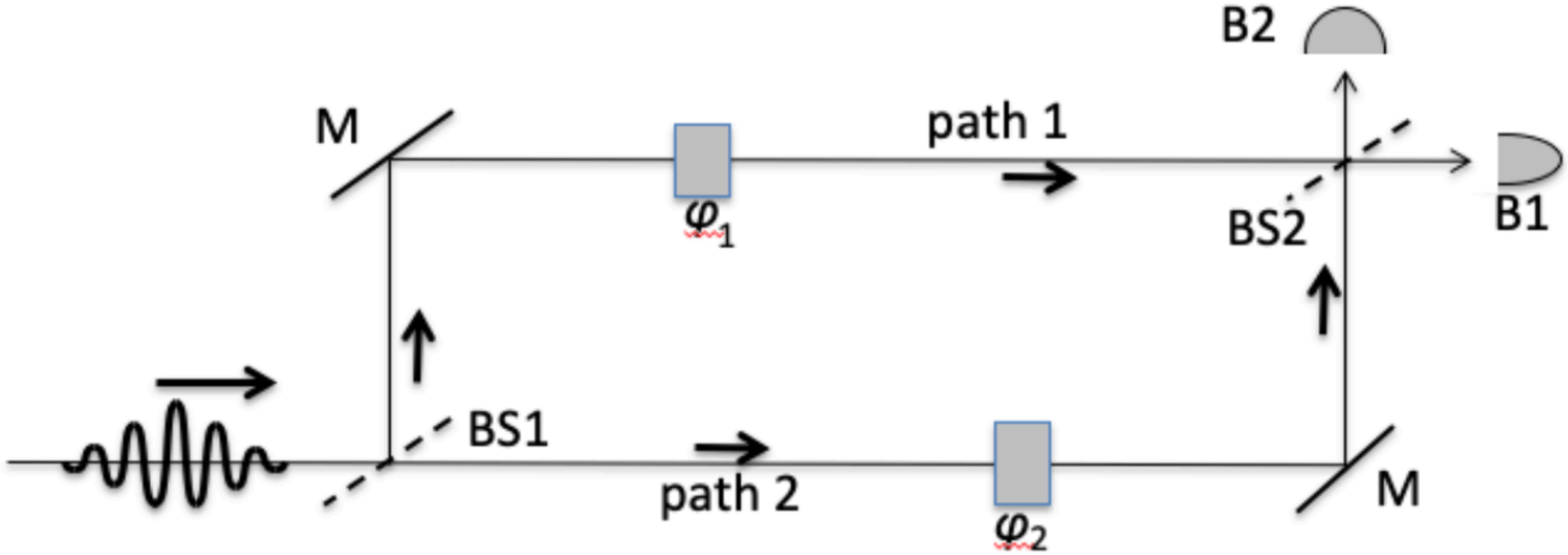


Figure 2. Mach-Zehnder interferometer experiment. A photon traverses a beam splitter, travels on two phase-shifted paths to another beam splitter, and is detected.

Figure 3 shows the results(discussed earlier).

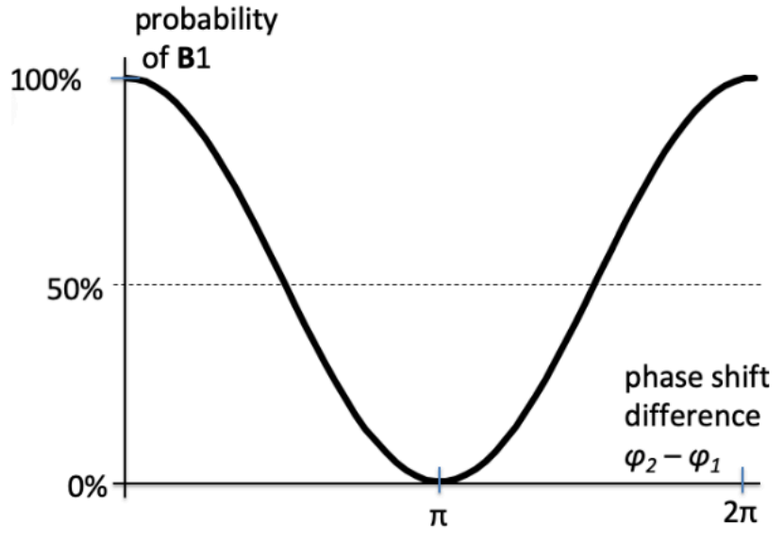


Figure 3. Single photon interference, pointing to Dirac's conclusion that "each photon ...interferes only with itself," i.e. each photon follows both paths.

Varying ϕ_1 through 180 degrees causes the photon's state to shift from 100% probability of detection at *B1*, through diminishing probabilities at *B1* and increasing probabilities at *B2*, finally reaching 100% probability of detection at *B2*.

The photon exhibits similar interference upon varying ϕ_2 .

Note that *A*'s state depends only on the phase *difference* $\phi_2 - \phi_1$.

Since single-trial results vary regardless of which phase shifter varies, it is hard to avoid the conclusion that each photon follows both paths.

In fact, let us assume the contrary, namely that each photon follows only one path.

Suppose the phase shifters are set to ensure 100% probability of detection at *B1*.

Under our assumption, this setting guarantees that every photon following path 1, and every photon following path 2, is detected at *B1*.

Suppose path 2 is now blocked without changing the phase settings, so that (still under our one-path assumption) every detected photon must now follow path 1 and be detected at *B1*.

But the experiment shows that, to the contrary, 50% of the detected photons now go to *B2*.

Conclusion: each photon follows both paths.

For a full discussion, see earlier lectures.

As Paul Dirac put it,

"The new theory, which connects the wave function with probabilities for one photon, gets over , the difficulty by making each photon go partly into each of the two components.

Each photon then interferes only with itself."

This illustrates why the "plus" sign in a superposition such as Equation (1) is interpreted by the word “AND".

We turn now to the entangled state Equation (2).

Many Bell inequality tests used polarization-entangled photon pairs to study the full phase dependence of this state.

More useful for this discussion are two interferometer experiments conducted in 1990.

Both of these experiments used *momentum*-entangled photon pairs to conduct Bell inequality tests of the entangled superposition Equation (2) (see earlier discussions).

Figure 4 shows the layout.

The source creates entangled pairs of photons *A* (moving leftward) and *B* (moving rightward) by laser down-conversion in a non-linear crystal.

The down-converted photons are prepared in the entirely microscopic state represented by $|\Psi_{AB}\rangle$ (Equation (2)) by selecting four single-photon beams, each a plane wave having a distinct momentum (i.e. wave vector), from the output of the crystal.

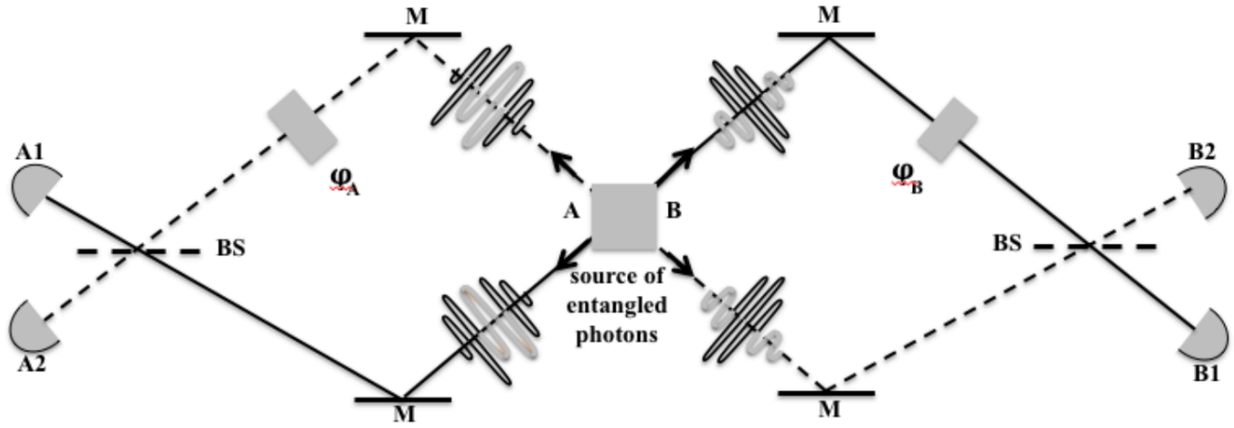


Figure 4. The RTO experiments. In each trial, each of two entangled photons travels two phase-shifted paths (one solid, the other dashed) to beam splitters and coincidence detectors. Think of one biphoton spreading from the source along both the solid and the dashed paths.

Figure 4 resembles two back-to-back Mach- Zehnder interferometer experiments (Figure 2) with BS1 located effectively inside the source.

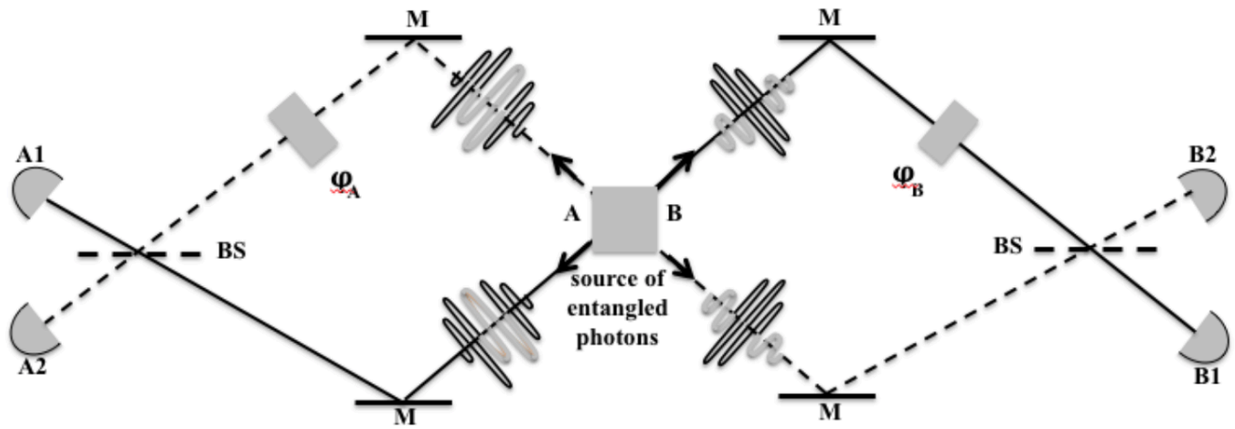


Figure 4. The RTO experiments. In each trial, each of two entangled photons travels two phase-shifted paths (one solid, the other dashed) to beam splitters and coincidence detectors. Think of one biphoton spreading from the source along both the solid and the dashed paths.

For simplicity and clarity, Figure 4 differs from the layout shown in the published papers.

In Figure 4, paired photons are directed oppositely.

This arrangement would result if the entanglement were prepared by the cascade decay of an atom.

In the actual experiments, however, down-converted photon pairs are emitted into two angular cones, resulting in photons that are not oppositely directed.

Figure 4's simpler geometry is pedagogically useful and has no effect on our arguments.

Although Figure 4 represents each photon as a wave packet spreading along two paths directed leftward or rightward, the composite system *AB* should be regarded as a *single* object, a **"biphoton".**

In each trial, a biphoton spreads outward from the source along two superposed branches.

One branch, represented by the first term $|A1\rangle\,|B1\rangle$ in Equation (2), spreads along the solid path and the other branch, $|A2\rangle\,|B2\rangle$, spreads along the dashed path.

As the biphoton *AB* moves outward along the solid path, *A* encounters a mirror *M,* then a beam splitter *BS* where it transmits and reflects to detectors *A1* and *A2*; photon *B* encounters a mirror *M,* a phase shifter ϕ_B, and a beam splitter *BS* where it transmits/reflects to detectors *B1*/*B2*.

The other half of the entanglement, namely the dashed path, has a similar description.

The experiments record outcomes at four photon detectors equipped with coincidence timers.

Published papers predict the experimental results theoretically and we **follow** their optical-path analysis here(slightly different but equivalent to our earlier arguments).

They begin by calculating the two-point nonlocal quantum field amplitudes $\Psi(Ai, Bj)$ at the four coincidence detectors *(Ai, Bj)*, and from these results they predict single-photon results.

For example, $\Psi(A1, B2)$ has two contributions, one from phase shifts in the beam following the solid path (the first term in $|\Psi_{AB}\rangle$) and the other from the dashed path (the second term).

From Equation (2), assuming distinct plane waves $\exp(i\mathbf{k}\cdot\mathbf{x})$ for each single-photon beam,

$$\Psi(A1, B2) = \frac{1}{2\sqrt{2}}\{\exp(i\phi_w)\exp[i(\phi_x + \phi_B)] + \exp[i(\phi_y + \phi^A)]\exp(i\phi_z)\} \tag{9}$$

where $\phi_w, \phi_x, \phi_y, \phi_z$ are fixed phase-shifts resulting from mirrors and beam splitters, and the additional factor of 1/2 comes from the superpositions created at the two beam splitters.

Using the Born rule, Equation (9) implies the coincidence probability

$$P(A1, B2) = |\Psi(A1, B2)|^2 = \frac{1}{4}[1 + \cos(\phi_B - \phi_A + \phi_v)] \tag{10}$$

where ϕ_u is a fixed phase arising from $\phi_w, \phi_x, \phi_y, \phi_z$. Similarly,

$$P(A1, B1) = \frac{1}{4}[1 + \cos(\phi_B - \phi_A + \phi_u)] \tag{11}$$

where ϕ_u is another fixed phase.

Remarkably, the sinusoidal terms predict coherent (phase-dependent) nonlocal interference between *A* and *B*, **regardless of their separation**.

There are similar expressions for *P(A2, B1)* and *P(A2, B2)*.

Single-photon predictions then follow. For example, from simple probability theory

$$P(A1) = P(A1, B1) + P(A1, B2)$$
$$= [1+\cos(\phi_B-\phi_A + \phi_u)]/4 + [1+\cos(\phi_B-\phi_A + \phi_v)]/4. \quad (12)$$

The theory papers then show the two fixed phase factors ϕ_u and ϕ_v differ by π:

$$\phi_v = \phi_u + \pi \ (mod\ 2\pi). \quad (13)$$

Thus the sinusoidal terms in Equation (12) *interfere destructively*, and *P(A1) = 1/2 regardless of phase*.

Equations (12) and (13) show this remarkable result to arise from destructive interference of two phase-dependent nonlocal contributions from the distant ***other*** photon *B*!

The result at all four single-photon detectors is the same:

$$P(A1) = P(A2) = P(B1) = P(B2) = 1/2. \quad (14)$$

Unlike the non-entangled single-photon superposition $|\Psi_A\rangle$, where the superposed photon is coherent (phase-dependent) as shown by Figure 3, each entangled photon is **"decohered"** and cannot interfere with itself.

Instead the two photons interfere with each other despite being separated by an arbitrary distance.

More accurately, ***each biphoton interferes with itself***.

Thus no single-photon interference fringes are associated with the state represented by $|\Psi_{AB}\rangle$.

Special relativity requires that this must be the case:

Since single-photon phase dependence could be used to establish an instantaneous communication channel between *A* and *B*, entanglement ***must*** deprive individual photons of their phase.

The result is **nonlocal coherence of the biphoton, AND decoherence of individual photons**.

The decoherence is required by special relativity.

Equation (14) can also be derived by tracing the pure state density operator $|\Psi_{AB}\rangle\langle\Psi_{AB}|$ over one subsystem to obtain the density operator for the other subsystem, which is what we did earlier.

This yields two density operators that appear to be mixtures but are not really "ignorance mixtures" as the word "mixture" is usually understood because the biphoton is in fact not in a mixed state but rather in a pure state represented by $|\Psi_{AB}\rangle$. That is why our earlier idea was incomplete!

The optical path analysis, above, derives Equation (14) while avoiding these controversial subtleties

A few definitions can put these predictions into more comprehensible form:

If one photon is detected in state 1 and the other in state 2, the two outcomes are said to be "different."

Otherwise, the outcomes are the "same."

Then from Equations (10) and (11), and similar expressions for *P(A2, B1)* and *P(A2, B2)*,

$$P(same) = P(A1,B1) + P(A2,B2) = 1/2[1 + cos(\phi_B - \phi_A)] \qquad (15)$$

$$P(different) = P(A1,B2) + P(A2,B1) = 1/2[1 - cos(\phi_B - \phi_A)]. \qquad (16)$$

Their difference, graphed in Figure 5, is called the "**degree of correlation**":

$$C = P(same) - P(different) = \cos(\phi_B - \phi_A). \tag{17}$$

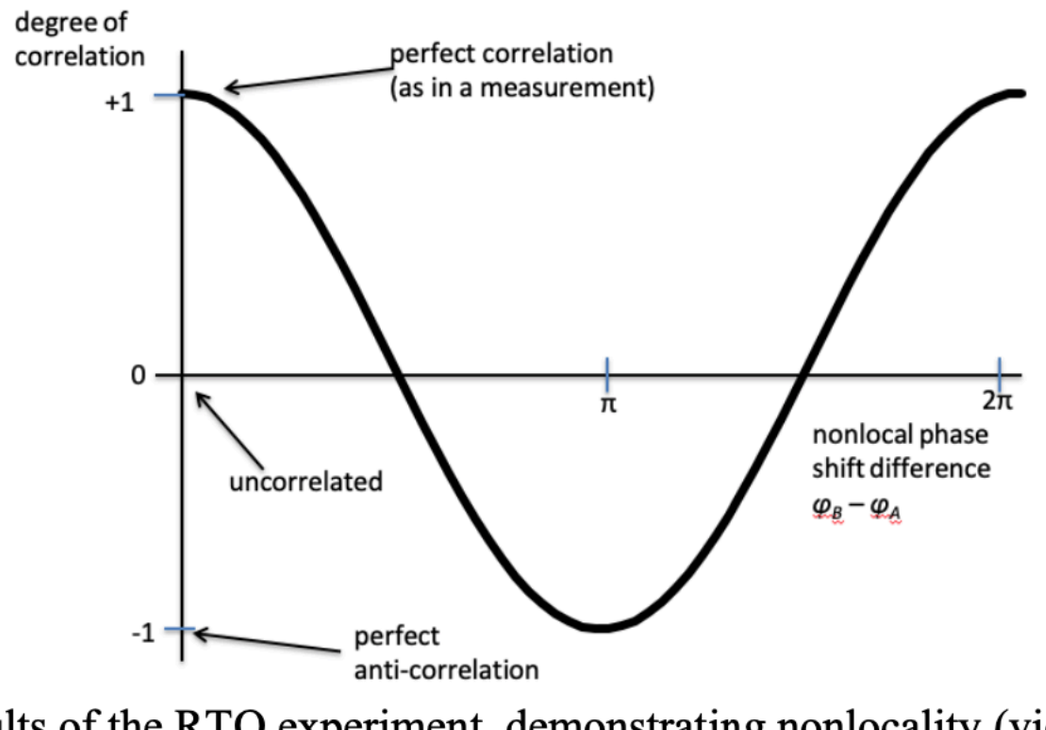


Figure 5. Results of the RTO experiment, demonstrating nonlocality (violation of Bell's inequality). The two photons interfere with each other across an arbitrary distance, i.e. each biphoton interferes with itself.

We now explore its physical significance.

5 - INTERPRETATION OF ENTANGLED MICROSCOPIC STATES

The original purpose of RTO's experiments was to demonstrate violations of Bell's inequality by comparing theoretical predictions, Figure 5, with experimental measurements.

The experimental results agreed with Figure 5 and violated Bell's inequality by 10 standard deviations, confirming the nonlocal nature of $|\Psi_{AB}\rangle$.

What does Figure 5 mean conceptually?

At zero phase difference, where the two phase shifters are set to equal phases, P(same) = 1 and C = P(same) - P(different) = +1.

Thus both stations always agree, despite the presence of beam splitters that randomize each photon prior to detection (see Figure 4).

It is as though coins were flipped at each station and they always came out either both heads or both tails!

Zero is the "measurement" phase angle where B's state is perfectly (and instantaneously) correlated with A's state.

The nonlocality is intuitively obvious:

Each photon acts like a detector of the state of the other photon regardless of separation!

Simply based on this conclusion, we can make an important observation about the entangled MS Equation (4):

Nonlocality is a central feature of quantum measurements.

For small non-zero phase differences, there is a small probability P(different) that results at the two stations will differ, i.e. observation of *B* no longer provides reliable information about *A*.

With increasing phase difference, this unreliability increases until, at $\pi/2$, the two detector pairs are entirely uncorrelated, and C = 0.

As the phase further increases from $\pi/2$ to π, P(different) increases while P(same) decreases, making C more and more negative.

Finally, C = -1 at phase difference π, implying perfect anti-correlation.

Thus *C* is aptly called the **"degree of correlation"**.

This description gives us a clear sense of the physical meaning of the fully entangled state Equation (2), indicating precisely which entities are superposed.

The biphoton's phase controls the degree to which the fixed phase-independent 50-50 states of its two spatially separated subsystems are ***statistically correlated***.

Compare this with the phase of the simple superposition $|\Psi_A\rangle$(Equation (1)), which controls the degree to which the single system *A* is represented by one or the other *state*.

The entities before and after the plus signs in Equations (1) and (2) are conceptually quite different:

Equation (1) sums two *states* while Equation (2) sums two *correlations between states*.

This **distinction is crucial**.

A ***state*** is a situation (or configuration or path) of a single quantum object, but a ***correlation*** is a statistical relationship between two (or more) quantum objects.

A ***superposition*** is the simultaneous existence of two or more states of a single quantum object.

An ***entanglement*** is the simultaneous existence of two or more relationships (specifically, correlations) between the states of two or more quantum objects.

Creating an entanglement is **quite a different matter** from creating a superposition.

Simple superposition of 1 photon.		Entangled superposition of 2 photons		
$\phi_2-\phi_1$	State of photon	$\phi_B-\phi_A$	State of each photon	Correlation between photons
0	100% 1, 0% 2	0	50-50 1 or 2	100% corr, 0% anticorr
$\pi/4$	71% 1, 29% 2	$\pi/4$	50-50 1 or 2	71% corr, 29% anticorr
$\pi/2$	50% 1, 50%2	$\pi/2$	50-50 1 or 2	50% corr, 50% anticorr
$3\pi/4$	29% 1, 71% 2	$3\pi/4$	50-50 1 or 2	29% corr, 71% anticorr
π	0% 1, 100% 2	π	50-50 1 or 2	0% corr, 100% anticorr

Table 1. Comparison between a simple superposition (Fig, 2) and an entangled superposition (Fig. 4). In Fig. 2, single-photon states vary with phase. In Fig. 4, only the *correlation between single-photon states* varies with phase while single-photon states have no phase. Thus *each biphoton is coherent* but its subsystems are incoherent. That is, entanglement *decoheres* each photon while transferring coherence to the biphoton.

To elaborate, Table 1 compares the superposition represented by $|\Psi_A\rangle$ (columns 1-2) with the entanglement represented by $|\Psi_{AB}\rangle$ (columns 3-5) at five different phases.

Column 2 demonstrates interference between the states represented by $|A1\rangle$ and $|A2\rangle$, implying the photon is in a superposition of following both paths and that the *state of A* varies with phase.

The phase dependence in column 2 shows that *A* interferes with itself.

In contrast, column 4 shows that, when the subsystems are represented by the pure state $|\Psi_{AB}\rangle$, neither photon has a phase.

Thus neither photon can interfere with itself, so **neither photon** can be represented by a superposition state.

They are decohered.

Both photons are **represented** by fixed, phase-independent, 50-50 states at all phase angles, **just as though they were in ignorance mixtures (which, however, they are not)**.

But phase dependence has not vanished, it has only been **transferred** to the composite system.

As column 5 reveals, the *degree of correlation between the fixed states of A and B* now varies with phase.

A photon represented by $|\Psi_A\rangle$ is in a coherent (phase-dependent) superposition of being in two states (i.e. of following two *paths)*.

$|\Psi_{AB}\rangle$, on the other hand, represents the coherent superposition of two *correlations between fixed states*.

Instead of two coherent *states* existing simultaneously, two coherent *relationships between states* exist simultaneously.

Neither subsystem is "smeared" (as Schrodinger apparently believed);

instead, *only the relationship between subsystems is smeared*.

Briefly,$|\Psi_A\rangle$ is a superposition of *states* and $|\Psi_{AB}\rangle$ is a superposition of *correlations*.

Thus $|\Psi_{AB}\rangle$ is qualitatively different from $|\Psi_A\rangle$.

$|\Psi_A\rangle$ exhibits properties of $|A1\rangle$ AND $|A2\rangle$, where "AND" indicates the superposition.

If you amplify *A* to macroscopic dimensions, you will get a macroscopic superposition.

$|\Psi_{AB}\rangle$ exhibits properties of correlations between $|A1\rangle$ and $|B1\rangle$ AND correlations between $|A2\rangle$ and $|B2\rangle$.

If you amplify *A* and *B* to macroscopic dimensions, you will *not* get a macroscopic superposition, you will simply get **correlations between macroscopic objects**.

The entanglement process transfers the coherence (phase dependence) of each photon to correlations between the two photons, leaving individual photons in mixtures that are incoherent but that are **not** ignorance mixtures.

$|\Psi_{AB}\rangle$ is a "superposition of correlations between properties," in contrast to $|\Psi_A\rangle$ which is a "superposition of properties."

There is a even better way to think about all this:

Regard *AB* as a single object, a biphoton.

Then Equation (2) describes a superposition of this object.

In the experiment (Figure 4), the two superposed states are represented by the solid line and the dashed line.

In this context, it makes no sense to speak of the superposition of a single subsystem, but it does make sense to speak of the superposition of the biphoton.

It is the biphoton that goes through the phases graphed in Figure 5 and indicated in Table 1 column 5.

Both branches (solid and dashed) of the biphoton exist simultaneously.

6 - INTERPRETATION OF THE MEASUREMENT STATE

Earlier we analyzed the microscopic state represented by Equation (2) mathematically, and then interpreted this state physically.

We now apply these insights to the entangled MS of a quantum system *A* and its detector *B* as derived in Equation (4).

In order for *B* to be a reliable detector, its states must be perfectly correlated with *A*'s states--it must exhibit $|Bi\rangle$ when and only when *A* is represented by $|Ai\rangle$ (*i = 1, 2*).

Thus Figure 5 implies the MS must be established at zero non-local phase: $\phi_B - \phi_A = 0$.

At this phase, two nonlocal perfect statistical correlations between a phase-independent 50-50 state of *A* and the corresponding phase-independent 50-50 state of *B* exist simultaneously.

As shown earlier, contrary to Schrodinger's description, *neither subsystem state can be "smeared out" (superposed) because neither subsystem has a phase*.

Instead, *correlations between* fixed states of *A* and *B* are smeared as shown in Table 1, while the detector indicates a single definite outcome.

Applying Table 1 to Schrodinger's example, the cat is predicted to be alive in 50% of trials, dead in the other 50%, and never in both states simultaneously.

Phase alterations would not smear the cat, they would smear only the correlations between the cat and the nucleus leading not to a smeared cat but **only** to imperfect detection.

There is no paradoxical macroscopic superposition.

But if neither *A* nor *B* is superposed, what *is* superposed?

What does the MS's "plus" sign really mean?

The answer, from Table 1 at zero phase:

$|A1\rangle$ is perfectly correlated with $|B1\rangle$ AND $|A2\rangle$ is perfectly correlated with $|B2\rangle$, where "AND" represents the superposition.

This simply says *both correlations exist simultaneously:*

$|A1\rangle$ if and only if $|B1\rangle$ AND $|A2\rangle$ if and only if $|B2\rangle$.

Again, there is no paradoxical macroscopic superposition. It's only the *correlations (relationships) between states*, not the states themselves, that are superposed.

Entanglement transforms a superposition of 2 *states* into a superposition of *two correlations between states*.

This makes quantum measurements possible because subsystem states can then be amplified to macroscopic dimensions **without requiring the creation of a macroscopic superposition**.

Neither subsystem is in a macroscopic superposition.

Since neither subsystem is superposed, only a single outcome occurs--a conclusion that also follows from Equation (14).

This single definite outcome occurs instantly upon entanglement, as facilitated by the nonlocal properties of the entangled MS.

Thus we have derived the collapse as an inevitable consequence of entanglement, and have **no need to postulate** such a process.

The MS **is** the collapsed state.

Our conclusion follows merely from standard principles of quantum theory with no other assumptions.

So von Neumann's enigmatic measurement state, Equation (4), is just what we want.

This entangled pure state provides the desired correlations, a single outcome, and the nonlocality required by Einstein's argument.

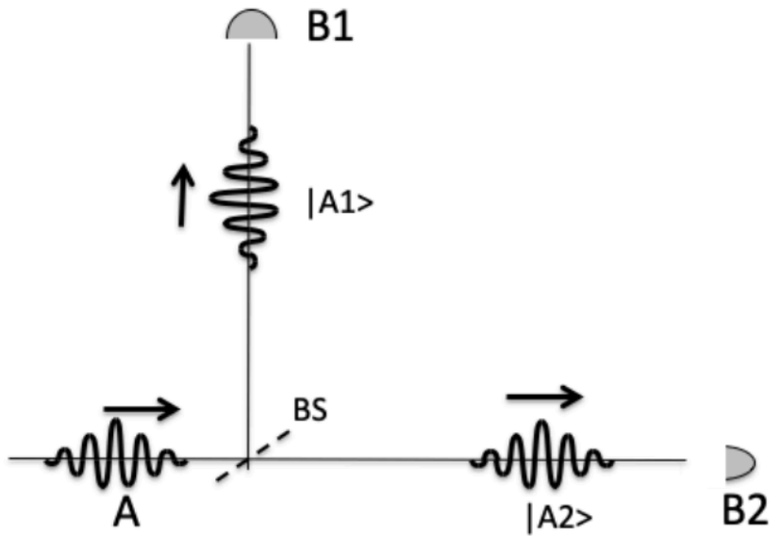
Note that the collapse is established at the microscopic level, prior to macroscopic amplification.

Next, we provide an example of the sequence of events.

7 - EXAMPLE

The following simple example typifies quantum measurements and illustrates the preceding insights in terms of a specific measurement process.

Consider the set-up in Figure 6.



A single photon traverses a beam splitter, creating the superposition represented by Equation (1) whose branches correspond to separate paths toward widely separated photon detectors.

Analogously to Figure 1, we assume the two detectors are equidistant from the beam splitter.

Each detector contains a photo-sensitive plate that, upon absorbing a photon, **releases** an electron.

Von Neumann's argument implies that, as the two branches of the superposition “approach” the detectors, at some point the branches overlap the detectors sufficiently that the entanglement process represented by Equation (4) occurs, where $|B_{ready}\rangle$ denotes the microscopic state of the detectors prior to entanglement while $|B1\rangle$ and $|B2\rangle$ denote their states following entanglement but prior to amplification and macroscopic recording.

At the instant of interaction, the state jumps from a superposition of two paths of *A* (Equation (1)) to a superposition of two correlations between *A* and *B* (right-hand side of the process in Equation (4)).

This entangled state is not paradoxical.

The right-hand side of Equation (4) entails precisely the proper correlations:

$$|A1\rangle \text{ if and only if } |B1\rangle \quad \text{AND} \quad |A2\rangle \text{ if and only if } |B2\rangle \ .$$

The excitation is transferred to only *one* detector while the other detector remains unexcited.

More correctly, either the solid branch or the dashed branch of the superposed biphoton (Figure 4) is randomly selected.

In fact, it has been shown experimentally and theoretically that the set-up shown in Figure 6 leads to entanglement and that the predicted nonlocal collapse occurs.

The nonlocality of the collapse is verified quantitatively by the experimental violation of a mathematical inequality derived based on QM.

Thus the collapse, a non-linear and irreversible process, occurs at the microscopic level.

Once one photoelectron is released, the process is thermodynamically irreversible because the electron is released into a vast number--a continuum--of free electron states and cannot feasibly be reversed.

This electron triggers an avalanche of other electrons leading to a macroscopic mark at one detector.
Other measurement set-ups follow the same general principles.

For example, in the measurement described by Einstein (discussed earlier), each small region of the detection screen acts as a single detector, and the diffracted electron's quantum state entangles with all these many regions.

Thus the argument above, involving only two detectors, must be extended to N detectors.

8 - SUMMARY AND CONCLUSION

Using only the standard principles of quantum physics, but minus the collapse postulate, we have shown that quantum state collapse occurs as a consequence of the entanglement that occurs upon measurement as described in 1932 by the von Neumann measurement state MS (Equation (4)).

The entangled "measurement state" of a quantum system and its detector is the collapsed state:

It incorporates the required perfect correlations between the system and its detector, it predicts precisely one definite outcome, and it incorporates the nonlocal properties--the instantaneous collapse across all branches of the superposition--that Einstein showed to be required in quantum measurements.

The measurement state Equation (4) does not describe a detector in a paradoxical superposition of displaying multiple outcomes, as had been supposed by Schrodinger and others.

Instead, quantum theory concludes that this state entails just what we expect following a measurement:

The states represented by $|A1\rangle$ and $|B1\rangle$ are perfectly correlated, AND the states represented by $|A2\rangle$ and $|B2\rangle$ are perfectly correlated, where "AND" represents the plus sign in the mathematical representation of the state.

Entanglement entails merely the simultaneous occurrence of two correlations between subsystems, not the simultaneous occurrence of two individual states of either subsystem.

There can be no paradoxical superposition of different detector states or of different system states, because the entanglement has shorn both the detector and the quantum system of their quantum phases - just as Feynmann supposed in the quote I gave earlier!

The phase has been transferred from the individual subsystems to the degree of correlation between subsystems.

To put all of this more directly, the single quantum object *AB* (the biphoton) collapses from a superposition to one of its members.

The measurement state's entanglement and its nonlocal properties, far from being paradoxical, are required in order to guarantee that the collapse occurs simultaneously across all branches of the superposition.

Eight previous insolubility proofs failed because they did not incorporate this required nonlocality.

Nonlocality is a central feature of quantum measurement.

There is no need for a special collapse postulate because the entangled state is the collapsed state.

Collapse occurs instantly upon entanglement.

This analysis should not be regarded as one more interpretation of quantum physics.

It is instead a **correction** of the previous misunderstanding of von Neumann's entangled measurement state.

It is not surprising that this misunderstanding has persisted for nearly 90 years.

After all, entanglement and nonlocality are deeply involved in the measurement problem's proper resolution but they only began to be understood in 1964, leading to a long period of gradual acceptance with confirmation only in 2015.

The delay in understanding measurement stemmed from this delay in understanding nonlocality.

For more than a century, much has been made of the odd and supposedly paradoxical nature of the quantum.

This presumed quantum spookiness has led to an excess of attempted fixes and interpretations.

Many experts have even declared the theory to be not a description of reality at all, but only a mathematical recipe that helps humans predict the results of experiments.

As Niels Bohr put it, **“There is no quantum world. There is only an abstract quantum description.”**

According to this hypothesis, quantum theory doesn’t describe anything real at all, so there’s no cause for concern about collapse of the quantum state and other odd quantum behaviors.

Such an easy resolution of the quantum quandaries amounts to giving up on science’s project of understanding the realities of the natural world.

It’s an extraordinary claim, requiring extraordinary proof. But there is no such proof, and there are no grounds for regarding quanta as any less real than rocks.

Indeed, rocks are made of quanta.

Although it has long been a conceit of humankind to imagine the universe to be centered around us, there is no reason to think that reality comprises only the kinds of things we experience in our own daily lives.

The real world does not fade from existence, nor does it become incomprehensible, at distances that happen to be several powers of ten smaller than teapots.

Atomic and subatomic processes are just as real as teapots and, with the help of technology, accessible to human experimentation and understanding.

From the viewpoint of the macroscopic and classical world that we are pleased to call “normal", there is certainly oddness in wave-particle duality, indeterminacy, quantum states, superposition, nonlocality, measurement, and quantum jumps.

But there are no logical contradictions here, no disagreements with experiment, and nothing that should persuade us that quantum physics is about anything other than the real world.

Quantum physics is either charmingly counterintuitive or maddeningly puzzling, depending on your taste, but it is entirely self-consistent and experimentally accurate.

It’s time to accept it with all its charms and puzzles, and stop trying to repair or reinterpret it.

It’s time, in other words, to relax and admit the world is not as we had thought.

Nature is far more creative than we could have conceived.

Our most fundamental theory is in better shape than its detractors suppose.

Quantum physics is a remarkable treasure trove of far-reaching phenomena and ideas whose surface we have probably only begun to scratch.

It’s time to fully embrace these ideas, incorporating them into our ways of thinking about the universe, about our planet, and about ourselves.

This is a process that will engage our minds and stretch our imaginations far into the future, for quantum physics is, indeed, not what anybody could have imagined.

Remembering one statement from Richard Feynman……

“Since we are not quantum objects we may never be able to see the details of this process, but it clearly happens. “

So let me just say.....

Don’t believe all the hype.

There is no need for many-worlds, multiverse and hidden variables, and so on.

There are no problems with standard quantum mechanics!

Now you have some idea of what I am thinking about and working on in the area of Foundations of QM!

Hope you had fun……..

If you want more details I could continue……..(but it gets more technical)