Advanced Quantum Mechanics

or

Watching a theoretical physicist at work as we attempt to build a strong physics/ mathematics background for our future efforts in Quantum Field Theory.

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 physical 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 - leaving your 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 will learn the necessary mathematics when needed.

Mostly, however, you will learn the mathematics as we use it in lecture and do problems.

Chapter 2

The Mathematics of Quantum Physics: Dirac Language

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

If I cover a topic 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 types of mappings are in the book.

Linear Vector Spaces

A vector space is defined with reference to a field(set on which addition, subtraction, multiplication, and division are defined).

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⟩ above 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

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

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

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

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 dependent 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, $k = 1, 2,\ldots, n$ 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.,

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

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

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

As we will see this is
Heisenberg's Uncertainty Principle

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 original 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, stated another way,

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

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

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 my 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 my 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:

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

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

Examples in my book and Problems.

Linear Operators

In general, an operator $\hat{Q}$ defined on a vector space **maps** vectors into vectors $: V \rightarrow V$, i.e., if $\hat{Q}$ = operator and $|k\rangle$ = a vector, then $\hat{Q}\,|k\rangle = |q\rangle$ = 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 $\hat{Q}\,|k\rangle$ is defined for all $|k\rangle$.

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 $|q_k\rangle$ 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

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

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

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

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

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

With matrices it corresponds to complex-conjugate-transpose of the matrix.

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

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

I hope you are already seeing that equations are much simpler when written in Dirac language!

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(much simpler), 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 my book for derivation

This 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

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

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

this is how you get

a 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

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

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

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

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

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

Called "spectral decomposition"

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.

Clearly, 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 my 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 $\hat{H}$

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

via the 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 my book and problems)

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

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

(details in my book and problems)

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 my book) in general

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

for any complete set of orthonormal vectors.

I emphasize that 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

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

Very important!!!

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 (see my book) 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, Spectral 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

As we will see, 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: First I will say it in "words" — — this way can be confusing…..

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

No confusions at all!!

That is why mathematics is the appropriate language!!

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

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

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 my 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*.

Another important result that can be derived is: Full derivation in my book.

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

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

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 many 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, sadly, do not attach much importance to Bayes' rule.

This rule, which relates prob(*A*|*B*∩*C*) to prob(*B*|*A*∩*C*), allows us to turn things around with respect to the conditioning symbol, which l**eads 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 the words **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 always 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.

My 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 $0 \leq n_A \leq n$.

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

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

(probability of A occurring)^r (probability of ~A occurring)$^{n-r}$

where

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

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

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

i.e., for n = 3, $n_A = 2$, we have possible orderings ppq, pqp, qpp

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

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

It is important to realize that 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 is in my book.

It is important to remember:

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

stop

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.

But consider this 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 “frequentist”

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.

For those interested…….

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

Also in my 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 additional 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.

We will proceed with Quantum Mechanics using the **standard frequency interpretation**.

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 possibly 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?

It is known 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.

End of Lecture #1

In what follows I will set up quantum mechanics using a new mathematical object - the **density operator**

I believe this approach, although more mathematical than more standard approaches, is the **best for understanding all the meanings associated with the mathematics**. and most clearly shows where all the **important physical features of quantum mechanics** come from

After we have the theory set, up we will use the developed **state vector idea** to work examples.

IMPORTANT: 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.

definition of Trace

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

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 discrete spectrum, we can write $\hat{W}$ in terms of its eigenvalues and eigenvectors (spectral representation) as

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

as we saw several times earlier

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

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

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

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

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

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

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 $\hat{B}$ we can always write

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

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

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

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

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, although all views end up with the same answers!

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. $\langle \hat{I} \rangle = 1$

Now to present Postulate #2.

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 will be 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 $|\alpha\rangle$, namely

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

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

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

We also have

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

and

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

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

If any operator satisfies this set of properties in a physical system, then that is all that is required for it to be defined(mathematically) as a density operator for that physical system.

Since,

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

we can then 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 example, 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

inserting $\hat{W}$

$$Tr(\hat{W}\hat{B}) = \sum_{m=1}^{2} \langle w_m | \hat{W}\hat{B} | w_m \rangle = \sum_{m=1}^{2} \langle w_m | \alpha \rangle \langle \alpha | \hat{B} | w_m \rangle$$

$$= \sum_{m=1}^{2} \langle w_m | w_2 \rangle \langle w_2 | \hat{B} | w_m \rangle = \sum_{m=1}^{2} \delta_{m2} \langle w_2 | \hat{B} | w_m \rangle$$

$$= \langle w_2 | \hat{B} | w_2 \rangle = \langle \alpha | \hat{B} | \alpha \rangle$$

Some extra details:

$\hat{W}|\beta\rangle = \hat{W}|w_1\rangle = w_1|w_1\rangle = 0$ since $w_1 = 0$

$\hat{W}|\alpha\rangle = \hat{W}|w_2\rangle = w_2|w_2\rangle = |w_2\rangle = |\alpha\rangle$ since $w_2 = 1$

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 $|\alpha\rangle$.

Thus, the density operator and the state vector are **equivalent** ways of representing a physical system **in this simple case of a pure state**.

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 $\hat{B}$ i.e., $\langle \hat{B} \rangle$, 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, as we stated earlier.

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 ⟶ The next step is to connect 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** of the pure state.

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

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

$\hat{W}$

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

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

Thus, the most important way of distinguishing whether a state is pure or not follows from this property of density operators, which says that:

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.

A very important result for measurement theory.

We will assume that both **pure** and **nonpure** (other texts use the names **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,

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

abbreviation = pdf

Now, the general definition of an average value using probability ideas 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

I do all derivations explicitly, but the detailed mathematics is only important if you intend to do complicated quantum calculations.

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 step function

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

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

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

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

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

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

do out all the algebra if you want to learn how to do it

Now we have

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

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

result from above inserted

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 result to be the case — — it is always important to check consistency among postulates.

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:

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

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

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

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

do out all the algebra if you want to learn how to do it

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 (they are degenerate) with eigenvalue $q = q_n$, we then have the general result that

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

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

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

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

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

do out all the algebra if you want to learn how to do it

If we have a pure state, then

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

and if q_n is a non-degenerate eigenvalue,

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

do out all the algebra if you want to learn how to do it

We note the following alternate and eventually 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

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

we just derived the Born Rule which is usually assumed in standard approaches to QM

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 or measuring that eigenvalue when in the state $|\alpha\rangle$, which is clearly given by

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

Thus, our result 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 my book.

Here we will **just assume** as in earlier derivations that variables are continuous and s**ums 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

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

do out all the algebra if you want to learn how to do it

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 (as earlier)

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

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.

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

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

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 $\hat{U}(\ell) = \hat{U}(\ell/2)\,\hat{U}(\ell/2)$. This says that

Remember linear $\quad \hat{A}[|a\rangle + |b\rangle] = \hat{A}|a\rangle + \hat{A}|b\rangle$

$\hat{A}c|a\rangle = c\hat{A}|a\rangle$

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.

Remember antilinear $\quad \hat{A}[|a\rangle + |b\rangle] = \hat{A}|a\rangle + \hat{A}|b\rangle$

$\hat{A}c|a\rangle = c^*\hat{A}|a\rangle$

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 $|q'_n\rangle = \hat{U}|q_n\rangle$ 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

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

We choose UNITARY operators because they do not change the NORM of vectors and we always want the vectors to have NORM =1.

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.

Algebra is simpler and results do not change

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

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

Taylor series expansion…..

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 then implies the condition to 1st order

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

If we let

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

then the unitarity 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 (all other variables held constant) 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

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

equation determining the unitary operator in terms of its associated generator

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, in this case.

This is just Stone's theorem, which is discussed in my 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

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

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

the least obvious rule

Wigner's theorem (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 shown, 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(a differential equation) 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 $\hat{H}$.

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 seem to be 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 also 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 my book, if interested.

Interaction Picture

We have derived the equation

now 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 , in general, have the simple solution

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

in this time-dependent 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 separate time developments, 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.

Most texts leave out this part and just **assume** expressions for $\hat{H}$.

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

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

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

written using matrices

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

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

or

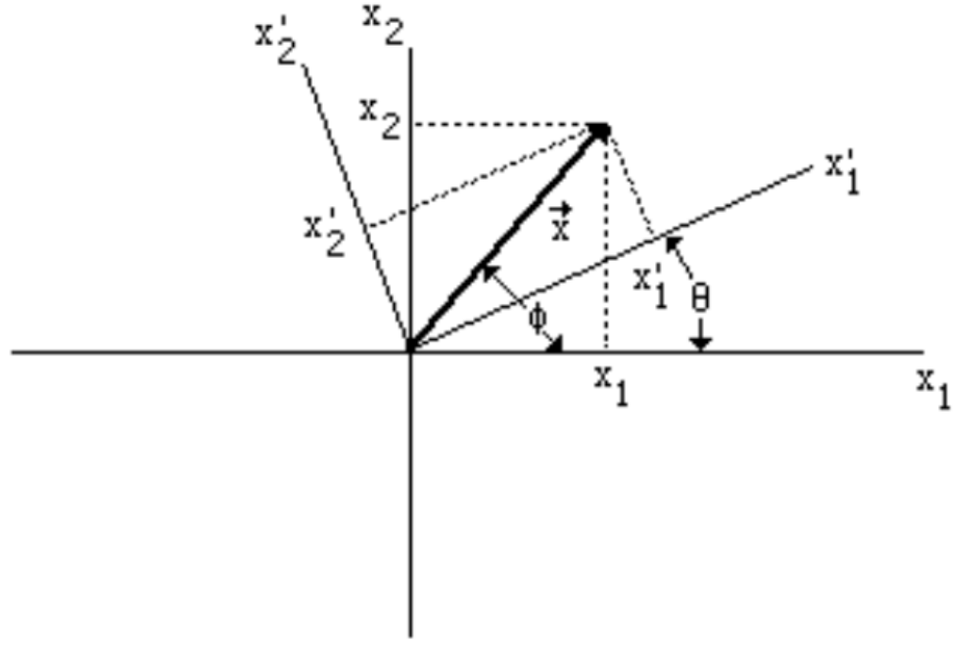
$$x'_1 = x_1 \cos\theta + x_2 \sin\theta$$

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

$$x'_3 = x_3$$

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:

presented for completness

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, as an example.

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 the most 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

$$|\psi\rangle \to |\psi'\rangle = \hat{U}(\tau)\,|\psi\rangle \text{ and } \hat{Q} \to \hat{Q}' = \hat{U}(\tau)\hat{Q}\hat{U}^{-1}(\tau)$$

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:

$$\text{if } |\psi(t)\rangle = \text{state vector at time } t, \text{ then for } t' = t + \tau$$

$$|\psi(t')\rangle = e^{i\tau \hat{H}} |\psi(t)\rangle$$

Generators of the Group Transformations

In general we can write

$$\hat{U}(s_\mu) = \prod_{\mu=1}^{10} e^{is_\mu \hat{K}_\mu}$$

contains 10 parameters

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 ,

$$\hat{U} = \hat{I} + i \sum_{\mu=1}^{10} s_\mu \hat{K}_\mu$$

Corresponds to only keeping terms LINEAR in the parameters.

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(after some algebra)

in my book

$$e^{i\epsilon \hat{K}_\mu} e^{i\epsilon \hat{K}_\nu} e^{-i\epsilon \hat{K}_\mu} e^{-i\epsilon \hat{K}_\nu} = \hat{I} + \epsilon^2 [\hat{K}_\nu, \hat{K}_\mu] + O(\epsilon^3)$$

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 (for those interested)below.

$$e^{i\epsilon\hat{K}_\mu} e^{i\epsilon\hat{K}_\nu} e^{-i\epsilon\hat{K}_\mu} e^{-i\epsilon\hat{K}_\nu}$$

$$= \left(\hat{I} + i\epsilon\hat{K}_\mu - \frac{1}{2}\epsilon^2\hat{K}_\mu^2\right)\left(\hat{I} + i\epsilon\hat{K}_\nu - \frac{1}{2}\epsilon^2\hat{K}_\nu^2\right) \times$$

$$\left(\hat{I} - i\epsilon\hat{K}_\mu - \frac{1}{2}\epsilon^2\hat{K}_\mu^2\right)\left(\hat{I} - i\epsilon\hat{K}_\nu - \frac{1}{2}\epsilon^2\hat{K}_\nu^2\right)$$

$$= \left(\hat{I} + i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2\hat{K}_\mu\hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)\right) \times$$

do the algebra

$$\left(\hat{I} - i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2\hat{K}_\mu\hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)\right)$$

$$= (\hat{I} + i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2\hat{K}_\mu\hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)$$

$$-i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2\hat{K}_\mu\hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)$$

$$+(i\epsilon(\hat{K}_\mu + \hat{K}_\nu))(-i\epsilon(\hat{K}_\mu + \hat{K}_\nu))$$

$$= \hat{I} - \epsilon^2\hat{K}_\mu^2 - \epsilon^2\hat{K}_\nu^2 - 2\epsilon^2\hat{K}_\mu\hat{K}_\nu + \epsilon^2\hat{K}_\mu^2 + \epsilon^2\hat{K}_\nu^2 + \epsilon^2\hat{K}_\mu\hat{K}_\nu + \epsilon^2\hat{K}_\nu\hat{K}_\mu$$

$$= \hat{I} + \epsilon^2\hat{K}_\nu\hat{K}_\mu - \epsilon^2\hat{K}_\mu\hat{K}_\nu$$

$$= \hat{I} + \epsilon^2[\hat{K}_\nu, \hat{K}_\mu] + O(\epsilon^3)$$

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 norms, i.e., $|\langle\phi|\psi\rangle|$).

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 generator commutators **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 (first-order) 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

stop

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 five unknown commutator types:

$$[\hat{J}_\alpha, \hat{P}_\beta], [\hat{G}_\alpha, \hat{H}], [\hat{J}_\alpha, \hat{J}_\beta], [\hat{G}_\alpha, \hat{J}_\beta], [\hat{G}_\alpha, \hat{P}_\beta]$$

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 $-\epsilon^2$ 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 $\epsilon_{\alpha\beta\gamma}$ 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 (we have been ignoring) has an effect on any of the commutators we have already determined.

In my 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** at this time.

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

$$\begin{aligned}\langle\psi(t)|\,\hat{\mathbf{V}}\,|\psi(t)\rangle &= \frac{d}{dt}\left(\langle\psi(t)|\,\hat{\mathbf{Q}}\,|\psi(t)\rangle\right)\\ &= \left[\frac{d}{dt}\langle\psi(t)|\right]\hat{\mathbf{Q}}\,|\psi(t)\rangle + \langle\psi(t)|\,\hat{\mathbf{Q}}\left[\frac{d}{dt}|\psi(t)\rangle\right]\end{aligned}$$

Now a time displacement corresponds to $t \longrightarrow t' = t + s$.

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.

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

$$i\frac{d}{du} e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}} = -e^{iu\hat{A}} \hat{A}\hat{B} e^{-iu\hat{A}} + e^{iu\hat{A}} \hat{B}\hat{A} e^{-iu\hat{A}}$$

$$= e^{iu\hat{A}} [\hat{A}, \hat{B}] e^{-iu\hat{A}} = [e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}}, \hat{A}]$$

very mathematical

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 and has been derived by these methods.

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 mathematical ideas (discussed earlier in detail in my book) that are needed to carry out the formal mathematics behind the next few derivation.

The examples that follow use these ideas to mathematically justify their results.

I state them below in their most general form for your reading if you wish……

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.

They involve the mathematics of subspaces and reduction of operators and sets.

At the mathematical level we are working, the following lecture examples only use these ideas as defined above without proof. We will look a Group Theory in all its glory later!!

We can proceed now….

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 —> meaning……

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(from example 0)

$$\hat{\mathbf{G}} - M\hat{\mathbf{Q}} = \text{ multiple of } \hat{I}$$

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

$$\hbar = \frac{1}{\beta}$$

we have inserted an unknown constant into the theory

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).

End of lecture #2

So far the examples have identified

$$\hat{G}_\alpha = M\hat{Q}_\alpha$$

$$\hat{\mathbf{P}} = M\hat{\mathbf{V}} \quad , \quad \hat{H} = \frac{1}{2}\hat{\mathbf{V}}\cdot\hat{\mathbf{V}} + E_0 \quad , \quad \hat{\mathbf{J}} = \hat{\mathbf{Q}}\times M\hat{\mathbf{V}}$$

$$[\hat{Q}_\alpha, \hat{p}_\beta] = i\hbar\delta_{\alpha\beta}\hat{I}$$

Two more examples will complete the picture.

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 $\hat{\mathbf{P}}$.

This means that an operator that commutes with this set $\{\hat{\mathbf{Q}}, \hat{\mathbf{P}}\}$ 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 greater 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 $\vec{a} = 0$.

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 that was used to get

$$\hat{H} = \frac{\hat{\mathbf{P}} \cdot \hat{\mathbf{P}}}{2M} + E_0\hat{I}$$

is still 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 $E_0 = c\ \hat{\mathbf{S}} \cdot \hat{\mathbf{S}}$ 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 new 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 $\hat{H}$.

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 $\hat{\mathbf{G}} = M\hat{\mathbf{Q}}$, it turns out that 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 $\hat{H}$.

So far 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 (all algebra for those interested)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.

I note that this entire development of symmetry operations was derived by Emmy Noether, whom I regard as one of the greatest mathematicians to have lived - she never really cared, however, about any applications!!!

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) My book next covers Multiparticle Systems, which we will skip over because we will not need this material for these lectures.

(2) My 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) My 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 $\hat{U}^\dagger(t_0,t_0)\hat{U}(t_0,t_0) = \hat{I}$ 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

$$\frac{d\hat{W}(t)}{dt} = \frac{d\hat{U}(t,t_0)}{dt}\hat{W}_0\hat{U}^\dagger(t,t_0) + \hat{U}(t,t_0)\hat{W}_0\frac{d\hat{U}^\dagger(t,t_0)}{dt}$$

$$= -\frac{i}{\hbar}\hat{H}\hat{U}(t,t_0)\hat{W}_0\hat{U}^\dagger(t,t_0) + \frac{i}{\hbar}\hat{U}(t,t_0)\hat{W}_0\hat{U}^\dagger(t,t_0)\hat{H}$$

$$= -\frac{i}{\hbar}[\hat{H}(t),\hat{W}(t)] \qquad \text{—-> equation of motion}$$

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. We now continue just using the Schrodinger picture.

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

$$\hat{A}\hat{T}(s) - \hat{T}(s)\hat{A} = [\hat{A}, \hat{T}(s)] = 0$$

commute with transformation operator

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

$$[\hat{K}, \hat{A}] = 0$$

commute with generator operator

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 transformations, 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.

This means that the 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 —> that is what appears as KET labels!

We will call the labels **quantum numbers**.

They will designate all we know about a state vector.

We now have enough tools to solve the quantum behavior of several real physical systems.

These are all systems that can be dealt without needing to use the Schrodinger differential equation, which we have not derived as yet.

It is the first indication, of many, that the normal Schrodinger equation is of limited use.

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

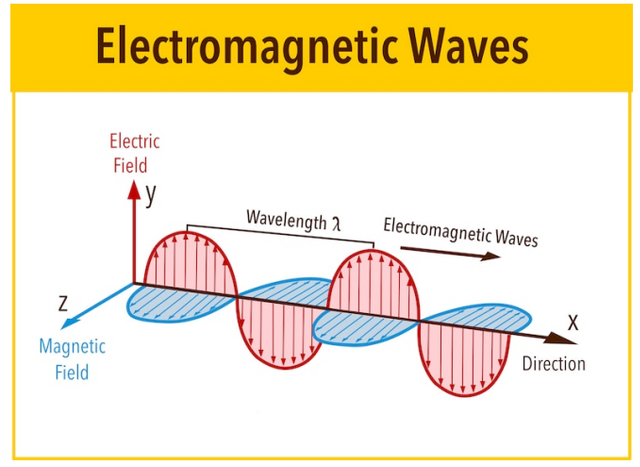
$$E = \frac{h}{2\pi}2\pi f = hf = h\frac{c}{\lambda} = pc$$

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 $\vec{E}$ vector .



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

$$E_x(\vec{r},t) = E_x^0 e^{i(kz-\omega t+\alpha_x)} \text{ and } E_y(\vec{r},t) = E_y^0 e^{i(kz-\omega t+\alpha_y)}$$

sines, cosines sines, cosines

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 $\hbar\,\omega$ (1 quantum).

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.

stop

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 class, 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 language, 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 (using matrix multiplication)

$$\langle\psi\,|\,\psi\rangle = \begin{pmatrix} \psi_x^* & \psi_y^* \end{pmatrix} \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix} = |\psi_x|^2 + |\psi_y|^2 = 1$$

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 our rules

$$\langle x \,|\, \psi \rangle = \langle x | \left(\psi_x |x\rangle + \psi_y |y\rangle \right) = \psi_x \langle x \,|\, x \rangle + \psi_y \langle x \,|\, y \rangle = \psi_x$$

$$\langle y \,|\, \psi \rangle = \langle y | \left(\psi_x |x\rangle + \psi_y |y\rangle \right) = \psi_x \langle y \,|\, x \rangle + \psi_y \langle y \,|\, y \rangle = \psi_y$$

or

$$|\psi\rangle = |x\rangle \langle x \,|\, \psi \rangle + |y\rangle \langle y \,|\, \psi \rangle$$

and similarly

spectral decompositions

$$|\psi\rangle = |R\rangle \langle R \,|\, \psi \rangle + |L\rangle \langle L \,|\, \psi \rangle$$

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?

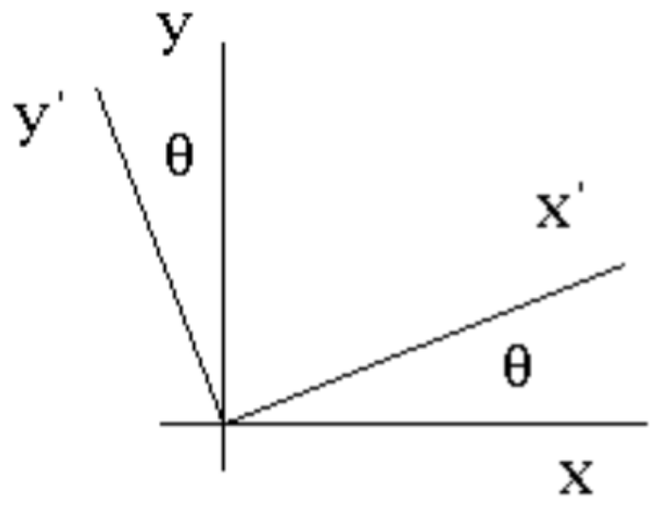
We have already seen two examples of basis sets for the 2-dimensional vector space of polarization states, namely,

$$\{|x\rangle, |y\rangle\} \text{ and } \{|R\rangle, |L\rangle\}$$

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

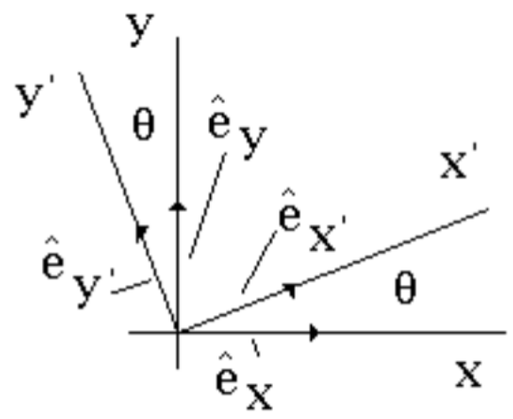
$$\begin{pmatrix}\langle x'|\psi\rangle\\ \langle y'|\psi\rangle\end{pmatrix} = \begin{pmatrix}\langle x'|x\rangle & \langle x'|y\rangle\\ \langle y'|x\rangle & \langle y'|y\rangle\end{pmatrix}\begin{pmatrix}\langle x|\psi\rangle\\ \langle y|\psi\rangle\end{pmatrix}$$

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 new 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

$$\begin{aligned}\hat{R}(\theta) &= \hat{R}(0) + \frac{1}{1!}\left.\frac{d\hat{R}(\theta)}{d\theta}\right|_{\theta=0}\theta + \frac{1}{2!}\left.\frac{d^2\hat{R}(\theta)}{d\theta^2}\right|_{\theta=0}\theta^2 \\ &\quad + \frac{1}{3!}\left.\frac{d^3\hat{R}(\theta)}{d\theta^3}\right|_{\theta=0}\theta^3 + \frac{1}{4!}\left.\frac{d^4\hat{R}(\theta)}{d\theta^4}\right|_{\theta=0}\theta^4 + \ldots \\ &= \hat{I} + \frac{1}{1!}\left(\frac{i\hat{J}_z}{\hbar}\right)\theta + \frac{1}{2!}\left(\frac{i\hat{J}_z}{\hbar}\right)^2\theta^2 + \frac{1}{3!}\left(\frac{i\hat{J}_z}{\hbar}\right)^3\theta^3 + \frac{1}{4!}\left(\frac{i\hat{J}_z}{\hbar}\right)^4\theta^4 + \ldots \\ &= e^{i\theta\hat{J}_z/\hbar}\end{aligned}$$

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

$$\begin{aligned}\hat{R}(\theta) = e^{i\theta\hat{J}_z/\hbar} &= \hat{I} + (i)\frac{\hat{J}_z}{\hbar}\theta + \frac{(i)^2\hat{I}}{2!}\theta^2 + (i)\frac{\hat{J}_z}{3!\hbar}\theta^3 + \frac{(i)^4\hat{I}}{4!}\theta^4 + \ldots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots\right) + i\left(\theta - \frac{\theta^3}{3!} + \ldots\right) \\ &= \cos\theta\hat{I} + i\sin\theta\hbar\hat{J}_z = \cos\theta\hat{I} + i\sin\theta\hat{Q}\end{aligned}$$

as stated 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

$$\hat{R}(\theta)|R\rangle = \left(\cos\theta\hat{I} + \frac{i}{\hbar}\sin\theta\hat{J}_z\right)|R\rangle$$
$$= (\cos\theta + i\sin\theta)|R\rangle$$
$$= e^{i\theta}|R\rangle$$

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

$$|\psi\rangle = |R\rangle\langle R|\psi\rangle + |L\rangle\langle L|\psi\rangle$$

use appropriate language

We then apply the rotation operator to obtain

$$\hat{R}(\theta)|\psi\rangle = \hat{R}(\theta)|R\rangle\langle R|\psi\rangle + \hat{R}(\theta)|L\rangle\langle L|\psi\rangle$$
$$= e^{i\theta}|R\rangle\langle R|\psi\rangle + e^{-i\theta}|L\rangle\langle L|\psi\rangle$$

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 momentum 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 polarization 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.

Start of Lecture 4

Amplitudes and Probabilities

Start of Lecture 4

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 added probabilities!!

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 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 using

$$|x\rangle = |R\rangle\langle R|x\rangle + |L\rangle\langle L|x\rangle$$

if we attempted to add a measurement to determine if the *x*-polarized photon was RCP or LCP, we would have 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 then 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)

$$\begin{aligned}
\langle \hat{A} \rangle = \langle x | \hat{A} | x \rangle = Tr(\hat{W}\hat{A}) &= \langle x | \hat{W}\hat{A} | x \rangle + \langle y | \hat{W}\hat{A} | y \rangle \\
&= \langle x | \hat{W}\hat{I}\hat{A} | x \rangle + \langle y | \hat{W}\hat{I}\hat{A} | y \rangle \\
&= \langle x | \hat{W} | x \rangle \langle x | \hat{A} | x \rangle + \langle x | \hat{W} | y \rangle \langle y | \hat{A} | x \rangle \\
&\quad + \langle y | \hat{W} | x \rangle \langle x | \hat{A} | y \rangle + \langle y | \hat{W} | y \rangle \langle y | \hat{A} | y \rangle
\end{aligned}$$

using

$$\hat{I} = |x\rangle \langle x| + |y\rangle \langle y|$$

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 the matrix for $\hat{W}$ is

$$\hat{W} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = |x\rangle \langle x|$$

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

$$\begin{aligned}\langle\hat{A}\rangle = \langle\psi|\hat{A}|\psi\rangle &= \frac{1}{2}\left[\langle x|\hat{A}|x\rangle + \langle x|\hat{A}|y\rangle + \langle y|\hat{A}|x\rangle + \langle y|\hat{A}|y\rangle\right] \\ &= Tr(\hat{W}\hat{A}) = \langle x|\hat{W}\hat{A}|x\rangle + \langle y|\hat{W}\hat{A}|y\rangle = \langle x|\hat{W}\hat{I}\hat{A}|x\rangle + \langle y|\hat{W}\hat{I}\hat{A}|y\rangle \\ &= \langle x|\hat{W}|x\rangle\langle x|\hat{A}|x\rangle + \langle x|\hat{W}|y\rangle\langle y|\hat{A}|x\rangle \\ &\quad + \langle y|\hat{W}|y\rangle\langle x|\hat{A}|x\rangle + \langle y|\hat{W}|y\rangle\langle y|\hat{A}|y\rangle\end{aligned}$$

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

$$\hat{W} = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1\\ 1\end{pmatrix}\begin{pmatrix}1 & 1\end{pmatrix} = |\psi\rangle\langle\psi|$$

if you do the algebra

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 $|a|^2 = |b|^2 = 1/2$.

We must choose

$$a = \frac{e^{i\alpha_a}}{\sqrt{2}} \text{ and } b = \frac{e^{i\alpha_b}}{\sqrt{2}}$$

Since we do not have any specific 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

$$\hat{W} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{2}|x\rangle\langle x| + \frac{1}{2}|y\rangle\langle y|$$
$$= Prob(x)|x\rangle\langle x| + Prob(y)|y\rangle\langle y|$$

note there are no cross-terms

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

$$\hat{U}_z = e^{ik_e z}\,|e\rangle\langle e| + e^{ik_o z}\,|o\rangle\langle o| \qquad \text{assuming motion in z-direction now}$$

is a *time development* operator of since ℓ = z = 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 clearly 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?

$$|\psi_{in}\rangle = a\,|0\rangle + b\,|e\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i\,|e\rangle)$$

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 we have developed.

stop

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 "the Energy operator"
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 to get a complete solution.

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 choose

$$\frac{\theta}{2} = \frac{B}{2A} \ll 1$$

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 ± 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, thus, 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 at the time.

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

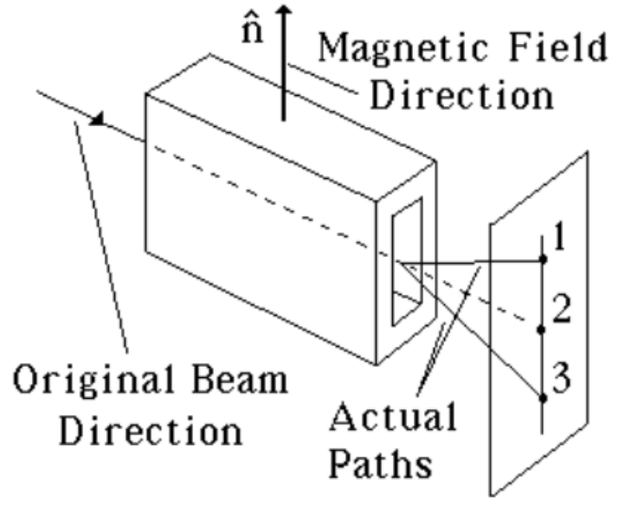
$$V = \vec{\mu} \cdot \vec{B}(\vec{r})$$

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 emerged from the oven had a random orientation.

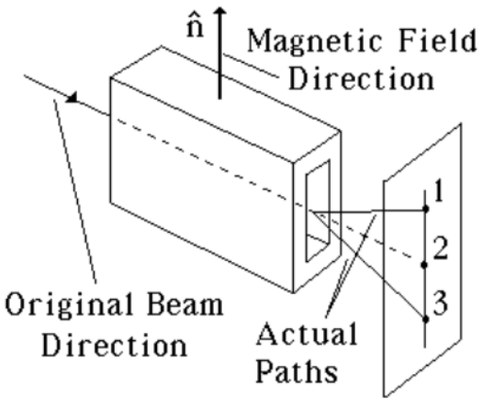
In general, a Stern-Gerlach apparatus takes an incoming beam of particles with angular momentum and splits the beam into a number of spots on a screen.

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).

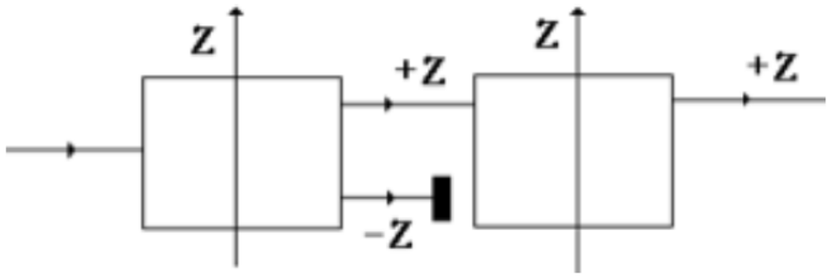
It contains $N/2$ particles.

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 get 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 $|+\ \hat{\mathbf{z}}\rangle$.

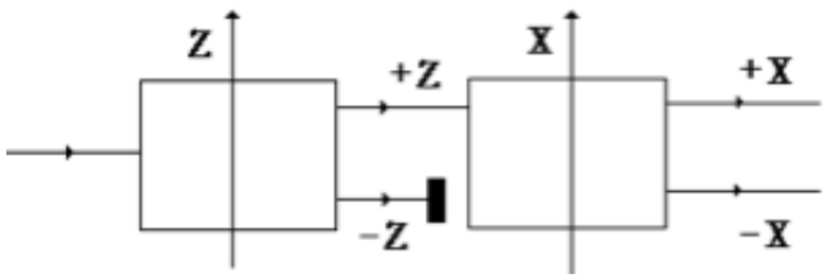
It contains $N/2$ particles.

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 $[\,\hat{J}_z, \hat{J}_x\,] \neq 0$.

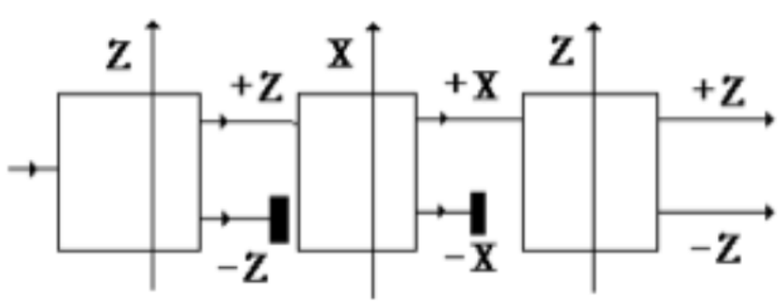
Experiment 3

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

$$P(\downarrow_z \mid \uparrow_x, \uparrow_z) = P(\downarrow_z \mid \uparrow_x) P(\uparrow_x \mid \uparrow_z) = |\langle \downarrow_z \mid \uparrow_x \rangle|^2 |\langle \uparrow_x \mid \uparrow_z \rangle|^2 = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

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!).

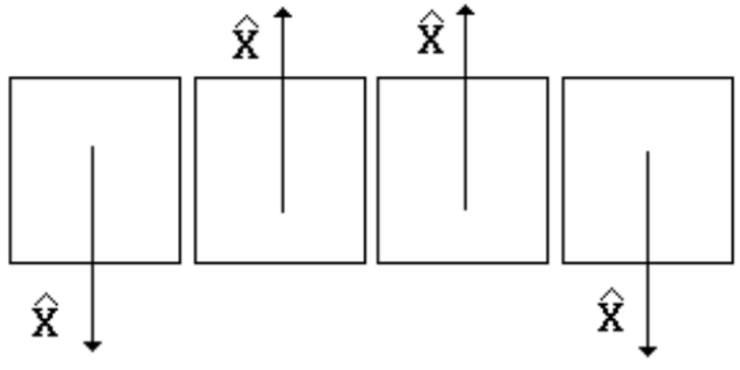
Again the **context** of the experiment is the significant feature!

Experiment 4

Now, let us construct a new device to illustrate this point.

It is called a **modified** SG$\hat{x}$ 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

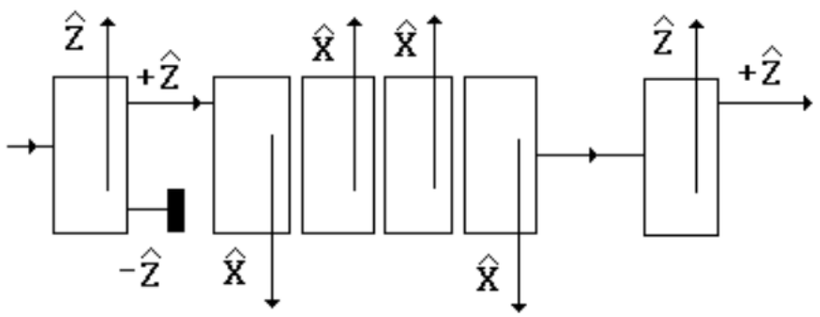
We can confirm this fact by inserting a modified SG$\hat{\mathbf{x}}$ device into Experiment 3 where we replace the SG$\hat{\mathbf{x}}$ device by the modified SG$\hat{\mathbf{x}}$ device.

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 $SG\hat{\mathbf{x}}$ 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

$$|+\hat{\mathbf{x}}\rangle = \frac{e^{1\alpha_+}}{\sqrt{2}}|+\hat{\mathbf{z}}\rangle + \frac{e^{1\alpha_+}}{\sqrt{2}}|-\hat{\mathbf{z}}\rangle$$

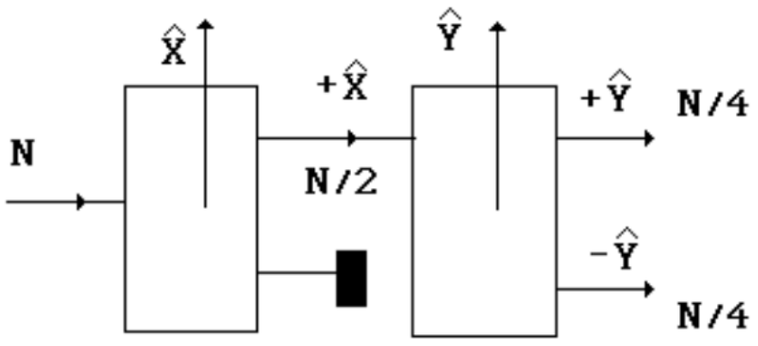
Similarly,

$$|-\hat{\mathbf{x}}\rangle = \frac{e^{1\beta_+}}{\sqrt{2}}|+\hat{\mathbf{z}}\rangle - \frac{e^{1\beta_+}}{\sqrt{2}}|-\hat{\mathbf{z}}\rangle$$

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.

next