8 The Heisenberg's Uncertainty Principle

1. **Simultaneous eigenstates**: Consider two operators that commute:

$$\left[\hat{A}, \hat{B}\right] = 0 \tag{8.0.1}$$

Let $\hat{A}$ satisfy the following eigenvalue equation:

$$\hat{A}\left|\eta\right\rangle = a\left|\eta\right\rangle \tag{8.0.2}$$

Multiplying both sides by $\hat{B}$

$$\hat{B}\left[\hat{A}\left|\eta\right\rangle\right] = \hat{B}\left[a\left|\eta\right\rangle\right] = a\hat{B}\left|\eta\right\rangle = a\left|\chi\right\rangle \tag{8.0.3}$$

where we have assumed $\hat{B}\left|\eta\right\rangle \equiv \left|\chi\right\rangle$.

Now lets try and use the commutation relation:

$$\hat{B}\left[\hat{A}\left|\eta\right\rangle\right] = \hat{A}\left[\hat{B}\left|\eta\right\rangle\right] = \hat{A}\left|\chi\right\rangle \tag{8.0.4}$$

From the right hand sides of the last two equations it follows that $\left|\chi\right\rangle$ is also an eigenvector of $\hat{A}$ with eigenvalue a. This could only happen if:

   (a) $\left|\chi\right\rangle = b\left|\eta\right\rangle$ since $\left|\eta\right\rangle$ is an eigenvector of $\hat{A}$ with eigenvalue a. Hence commuting operators have *simultaneous* eigenstates. That is these can be exactly measured simultaneously. (By extension two operators that do not commute cannot be measured simultaneously as we will see in the next section.) This is a very important distinguishing factor with respect to classical mechanics. In classical mechanics you can measure any two observables simultaneously. In quantum mechanics, only variables whose (Hermitian) operators commute can be observed simultaneously. Now go back and read the Stern-Gerlach experiment. It must mean that since S_z, S_x and S_y cannot be observed simultaneously, *the operators corresponding to these observables, the Pauli spin matrices, do not commute*.

   (b) There is a second possibility and that is $\left|\chi\right\rangle$ is not equal to a constant times $\left|\eta\right\rangle$. In this case the operator $\hat{A}$ must have degenerate states ($\left|\chi\right\rangle$ and $\left|\eta\right\rangle$) and the operator $\hat{B}$ can be used to obtain all the degenerate states of $\hat{A}$. But even in this case, the non-degenerate eigenvectors of $\hat{A}$ are *simultaneously* eigenstates of $\hat{B}$ as is clear from the previous case.

2. Actually case (b) is also deeply connected with Stern-Gerlach as we will see later.

3. **Expectation or Average value of an operator**

If the $\left|\psi\right\rangle$ represents the state of a system, the *expectation value* of an operator $\hat{A}$ with respect to that state is given by the quantity

$$\left\langle\hat{A}\right\rangle = \left\langle\psi\right|\hat{A}\left|\psi\right\rangle \tag{8.0.5}$$

What does this mean? If there is an observable associated with the operator $\hat{A}$, then $\left\langle \hat{A} \right\rangle$ is **the average measured value of the observable.**

(a) Case 1: The state vector ψ is an eigenstate of the operator $\hat{A}$. In that case, if a is the associated eigenvalue then $\left\langle \hat{A} \right\rangle = \langle\psi| \hat{A} |\psi\rangle = a$. (Note: $\langle\psi|\, \psi\rangle = 1$.) That is the observable has a measured value equal to the eigenvalue of the operator $\hat{A}$. Consider Figure 6 (reproduced below) as a way to understand this concept. Here, an S_z^+ state is sent into an SGz measuring device and the resultant measurement is the eigenvalue of the S_z^+ state with respect to the $\hat{S}_z$ operator. Hence if ψ is an eigenstate of the operator, the corresponding measured value, or expectation value is a,

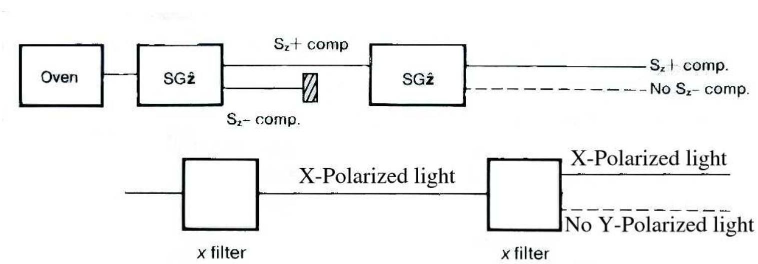


Figure 19:

(b) Case 2: The state vector ψ is not an eigenstate of the operator $\hat{A}$. In this case, if $\hat{A}$ is a Hermitian operator then the *eigenstates of a Hermitian operator form a complete ortho-normal set*. **Hence any vector like ψ can be expanded as a linear combination of such a complete set of vectors.** Therefore, if $\{\chi_i\}$ are the family of eigenvectors of $\hat{A}$, with eigenvalues $\{a_i\}$, that is

$$\hat{A}\chi_i = a_i\chi_i \qquad i = 1, 2, 3, \cdots \tag{8.0.6}$$

then we can expand the vector ψ as a linear combination of the vectors $\{\chi_i\}$, as

$$|\psi\rangle = \left[\sum_i |\chi_i\rangle \langle\chi_i|\right] |\psi\rangle = \sum_i |\chi_i\rangle \langle\chi_i| \, |\psi\rangle \tag{8.0.7}$$

Note we have used the "resolution of identity" here. If we denote the number $\langle\chi_i| \, |\psi\rangle = c_i$ (note that this is a "dot" product and hence is a number), then

$$|\psi\rangle = \sum_i c_i |\chi_i\rangle \tag{8.0.8}$$

The expectation value of the operator $\hat{A}$ can then be written as

$$\left\langle \hat{A} \right\rangle = \langle\psi| \hat{A} |\psi\rangle \quad = \quad \left\{\sum_j c_j{}^* \langle\chi_j|\right\} \hat{A} \left\{\sum_i c_i |\chi_i\rangle\right\}$$

$$
\begin{aligned}
&= \sum_j \sum_i \left[{c_j}^* \langle \chi_j | \right] \hat{A} \left[c_i | \chi_i \rangle \right] \\
&= \sum_j \sum_i {c_j}^* c_i \langle \chi_j | \hat{A} | \chi_i \rangle
\end{aligned} \tag{8.0.9}
$$

Note that $\hat{A} |\chi_i\rangle = a_i |\chi_i\rangle$ and since $\langle \chi_j | |\chi_i\rangle = \delta_{i,j}$ (since the eigenfunctions of $\hat{A}$ are orthonormal!!) we can simplify the above equation to

$$
\begin{aligned}
\left\langle \hat{A} \right\rangle &= \sum_j \sum_i {c_j}^* c_i a_j \langle \chi_j | |\chi_i\rangle \\
&= \sum_j \sum_i {c_j}^* c_i a_i \delta_{i,j}
\end{aligned} \tag{8.0.10}
$$

The last term has two summations. The summation over j yields a non-zero value onlu when $j = i$ on account of the Kronecker delta: $\delta_{i,j}$. Hence we have

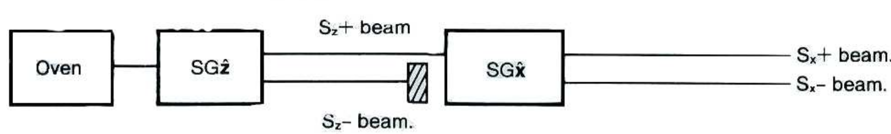
$$
\left\langle \hat{A} \right\rangle = \sum_i {c_i}^* c_i a_i = \sum_i N_i a_i \tag{8.0.11}
$$

where we have just substituted $N_i = {c_i}^* c_i$.

Now the last equation does look like an average of all the a_i values. (Why? Note that ${c_i}^* c_i \leq 1$ because ψ should be normalized and that implies $\sum_i {c_i}^* c_i = 1$ from Eq. (8.0.8).)

When you perform a large number of measurements, the average of all those measurements will be $\left\langle \hat{A} \right\rangle$. That is why it is called the "expectation value". But a single measurement will yield a value corresponding to one of the eigenstates of $\hat{A}$!!! The average of all these measurements, just like in Eq. (8.0.11), will give $\left\langle \hat{A} \right\rangle$.

Consider the SG experiments and see if you can rationalize this fact. Especially consider the experiment



You performed this experiment using the java script. Lets now look at the average value of S_x after the second measurement. When you conduct a large number of experiments (say 10,000 runs), you get an almost equal amount of S_x^+ and S_x^- and the average or expectation value turns out to be zero. (We rationalized based on the projection of the S_z^+ state on S_x^+ and S_x^- that indeed we should get equal amounts of both.)

However, when you perform just one experiment you get either S_x^+ or S_x^-. Hence, the statement above, a single measurement will yield a value corresponding to one of the eigenstates of $\hat{A}$!!! The average of all these measurements, just like in Eq. (8.0.11), will give $\left\langle \hat{A} \right\rangle$.

Now we are ready to find out what the Heisenberg's Uncertainty Principle really is, in all its glory. There are two slightly different ways to derive this and we will study both ways. The first approach is rigorous and thats useful. The second is more elegant and leads to fun things such as coherent states. (Incidentally Glauber got the Nobel prize for physics in 2005, for his work on coherent states.)

Let $\hat{A}$ and $\hat{B}$ be two Hermitian operators. Lets define the commutator of these operators as:

$$\left[\hat{A}, \hat{B}\right] = \imath\hat{C} \tag{8.0.12}$$

where $\hat{C}$ is another operator.

Homework: Prove that $\hat{C}$ is a Hermitian operator. (To prove this you will need to know that $\left[\hat{A}\hat{B}\right]^{\dagger} = \hat{B}^{\dagger}\hat{A}^{\dagger}$.)

Now lets define an operator $\hat{D}$ that is a complex linear combination of $\hat{A}$ and $\hat{B}$:

$$\hat{D} = \hat{A} + (\alpha + \imath\beta)\,\hat{B} \tag{8.0.13}$$

$\hat{D}$ is a complex linear combinations of $\hat{A}$ and $\hat{B}$. Note that $\alpha + \imath\beta$ is a complex number.

Now $\hat{D}$ is an operator and hence we can say:

$$\hat{D}\,|S\rangle = |Q\rangle \tag{8.0.14}$$

That is the operator $\hat{D}$ *contains a recipe* that converts the *ket* vector $|S\rangle$ to the *ket* vector $|Q\rangle$

Now, the "dot" product of $|Q\rangle$ with itself, that is the quantity $\langle Q\,|Q\rangle \geq 0$. (Why? The dot product of any two vectors is always greater that or equal to zero. Remind yourself that this is indeed true by going back to Section 2.8.1 and the appendix.)

This means:

$$\langle Q\,|Q\rangle = \langle S|\,\hat{D}^{\dagger}\hat{D}\,|S\rangle \geq 0 \tag{8.0.15}$$

Note, we have used: $\langle S|\,\hat{D}^{\dagger} \equiv \langle Q|$.

If we use Eq. (8.0.13) in Eq. (8.0.15):

$$\langle S|\left\{\hat{A} + \alpha\hat{B} + \imath\beta\hat{B}\right\}^{\dagger}\left\{\hat{A} + \alpha\hat{B} + \imath\beta\hat{B}\right\}|S\rangle \geq 0 \tag{8.0.16}$$

Since $\hat{A}^{\dagger} = \hat{A}$ and $\hat{B}^{\dagger} = \hat{B}$ since they are Hermitian, this implies

$$\left\langle \hat{A}^2 \right\rangle + \left(\alpha^2 + \beta^2\right)\left\langle \hat{B}^2 \right\rangle + \alpha\left\langle \hat{C}' \right\rangle - \beta\left\langle \hat{C} \right\rangle \geq 0 \tag{8.0.17}$$

where $\hat{C}' = \left[\hat{A}, \hat{B}\right]_{+}$. Now we can rewrite the left side in the above equation in the following fashion:

$$\begin{aligned} \left\langle \hat{A}^2 \right\rangle &+ \left\langle \hat{B}^2 \right\rangle \left(\alpha + \frac{1}{2}\frac{\left\langle \hat{C}' \right\rangle}{\left\langle \hat{B}^2 \right\rangle}\right)^2 + \left\langle \hat{B}^2 \right\rangle \left(\beta - \frac{1}{2}\frac{\left\langle \hat{C} \right\rangle}{\left\langle \hat{B}^2 \right\rangle}\right)^2 \\ &- \frac{1}{4}\frac{\left\langle \hat{C}' \right\rangle^2}{\left\langle \hat{B}^2 \right\rangle} - \frac{1}{4}\frac{\left\langle \hat{C} \right\rangle^2}{\left\langle \hat{B}^2 \right\rangle} \geq 0 \end{aligned} \tag{8.0.18}$$

Since the above expression holds for all α and β we are certainly free to choose the value of these variables as per our convenience. In particular we choose these variables to simplify our algebra. We choose α and β such that the bracketed terms become zero, leading to:

$$\left\langle \hat{A}^2 \right\rangle - \frac{1}{4}\frac{\left\langle \hat{C}' \right\rangle^2}{\left\langle \hat{B}^2 \right\rangle} - \frac{1}{4}\frac{\left\langle \hat{C} \right\rangle^2}{\left\langle \hat{B}^2 \right\rangle} \geq 0 \tag{8.0.19}$$

Multiplying by $\left\langle \hat{B}^2 \right\rangle$, we obtain:

$$\left\langle \hat{A}^2 \right\rangle \left\langle \hat{B}^2 \right\rangle - \frac{1}{4}\left(\left\langle \hat{C}' \right\rangle^2 + \left\langle \hat{C} \right\rangle^2 \right) \geq 0 \tag{8.0.20}$$

That is,

$$\left\langle \hat{A}^2 \right\rangle \left\langle \hat{B}^2 \right\rangle \geq \frac{1}{4}\left(\left\langle \hat{C}' \right\rangle^2 + \left\langle \hat{C} \right\rangle^2 \right) \geq \frac{1}{4}\left\langle \hat{C} \right\rangle^2 \tag{8.0.21}$$

Note, we have used the fact that $\left\langle \hat{C}' \right\rangle^2 \geq 0$, the expectation value of a Hermitian operator is always greater than zero.

Homework: Prove the $\hat{C}'$ is Hermitian operator. Also show that $\left\langle \hat{C}' \right\rangle^2 \geq 0$.

Now the uncertainty in the observable A is defined as:

$$\hat{\Delta A} = \hat{A} - \left\langle \hat{A} \right\rangle \tag{8.0.22}$$

Does it make sense that this is the uncertainty in A? The second term on the right side is the average measured value or expectation value. The action of the first term on a ket yields a eigenstate of A which could in general be different from the expectation value. (Remember the SG example from the discussion expectation values.)

This implies the average uncertainly is given by

$$\left\langle \left(\hat{\Delta A}\right)^2 \right\rangle \equiv (\Delta A)^2 = \left\langle \left(\hat{A} - \left\langle \hat{A} \right\rangle\right)^2 \right\rangle = \left\langle \hat{A}^2 \right\rangle - \left\langle \hat{A} \right\rangle^2 \tag{8.0.23}$$

And similarly for the operator $\hat{B}$ we have:

$$\left\langle \left(\hat{\Delta B}\right)^2 \right\rangle \equiv (\Delta B)^2 = \left\langle \left(\hat{B} - \left\langle \hat{B} \right\rangle\right)^2 \right\rangle = \left\langle \hat{B}^2 \right\rangle - \left\langle \hat{B} \right\rangle^2 \tag{8.0.24}$$

If we assume the average expectation values $\left\langle \hat{A} \right\rangle = \left\langle \hat{B} \right\rangle = 0$,

$$(\Delta A)^2 (\Delta B)^2 \geq \frac{1}{4}\left\langle \hat{C} \right\rangle^2 \tag{8.0.25}$$

or

$$\Delta A \Delta B \geq \frac{1}{2}\left| \left\langle \hat{C} \right\rangle \right| \tag{8.0.26}$$

Equation (8.0.26) is called the Heisenberg uncertainty principle.

This equation suggests that one cannot specify, simultaneously, exact values (eigenvalues) of a pair of non-commuting observables (e.g., position and momentum as we will see further down below) and places quantitative restrictions on their relative variances.

Now, ΔA and ΔB uncertainties in a measurement of A and B. The equation above implies that if the operators do not commute they cannot be simultaneously meassured with infinite certainty. *Remember we learnt earlier that commuting operators simultaneous eigenstates. When they do not commute, their eigenstates may be different leading to the fact that they cannot be simultaneously observed and hence the uncertainly principle above.*

The essential origin of this principle is that quantum mechanics possesses the mathematical structure of a linear vector space (viz., a Hilbert space). *Note we have implicitly used nothing but vector spaces to derive uncertainty. Why? All we assumed is operators act on kets, and yield new kets.*

1. **Uncertainty of position and momentum.** You may recall based on your homework that:

$$[\hat{x}, \hat{p}] = \left[\hat{x}, -\imath\hbar\frac{\partial}{\partial x}\right] = \imath\hbar \tag{8.0.27}$$

Thus using this in Eq. (8.0.26) we obtain:

$$\Delta x \Delta p \geq \frac{1}{2}\hbar \tag{8.0.28}$$

Thus one cannot specify, simultaneously, the position and momentum of a system. As stated earlier this entirely due to the mathematical structure of a linear vector space that we have been forced to introduce on account of the Stern-Gerlach experiments!! The transformation between the position representation of the vector space and the momentum representation is the Fourier transform as we have seen earlier, in Eq. (**??**), the collection of waves or wave-packet. ($\exp[\imath kx]$ is the momentum eigenstate.)

2. **The Mimimum Uncertainty Wavepackets and Coherent States**

There is another way to derive the Heisenberg's uncertainty principle. We will consider that here briefly to expose a very important concept. You may wonder Eqs. (8.0.26) and (8.0.28) are inequalities. Are there conditions when the equality is valid? That is to put the question a different way, what are the functions, or states, that have the *minimum uncertainty product* $\Delta x\Delta p = \frac{1}{2}\hbar$. We will see here that these functions are basically a "Gaussian multiplied by a plane-wave" (or moving gaussians since the plane wave being an eigenstate of momentum simply translates the gaussian) and are called as the *minimum uncertainly wavepackets" or "coherent states" more commonly.*

Reconsider Eq. (8.0.23) in the following form:

$$\begin{aligned}\left\langle \left(\hat{\Delta A}\right)^2 \right\rangle \equiv \left\langle \left(\hat{A} - \left\langle \hat{A} \right\rangle\right)^2 \right\rangle &= \left\langle f_1 \middle| \left(\hat{A} - \left\langle \hat{A} \right\rangle\right)^\dagger \left(\hat{A} - \left\langle \hat{A} \right\rangle\right) \middle| f_1 \right\rangle \\ &= \langle g_1 | \, g_1 \rangle\end{aligned}$$

$$
\begin{aligned}
&= \int dx g_1^*(x) g_1(x) \\
&= \int dx |g_1(x)|^2 \qquad (8.0.29)
\end{aligned}
$$

where $|g_1\rangle = \left(\hat{A} - \left\langle \hat{A} \right\rangle\right) |f_1\rangle$ and hence $\langle g_1| = \langle f_1| \left(\hat{A} - \left\langle \hat{A} \right\rangle\right)^\dagger$

In the last part of Eq. (8.0.29) we have resolved the identity using the coordinate representation. (Remember what that means? Resolution of the identity, that is 1, is inserted in the middle in $\langle g_1|\ g_1\rangle$.)

Now, we will invoke a useful mathematical tool called the Schwartz inequality which says that for any two functions g_1 and g_2:

$$
\left(\int dx |g_1(x)|^2\right)\left(\int dx |g_2(x)|^2\right) \geq \left|\int dx g_1(x) * g_2(x)\right|^2 \qquad (8.0.30)
$$

So as not to lose the flow we will accept this identity for now, but you can rationalize this identity by considering the fact that the two terms on the left are "dot" products of g_1 with itself, and g_2 with itself. The equality in Eq. (8.0.30) only holds when $g_2 \propto g_1$.

Since a similar equality like Eq. (8.0.29) holds for operator $\hat{B}$ (that is replace $\hat{A}$ with $\hat{B}$ in Eq. (8.0.29) and using the function $|g_2\rangle = \left(\hat{B} - \left\langle \hat{B} \right\rangle\right) |f_1\rangle$), we can say:

$$
\left\langle \left(\Delta \hat{A}\right)^2 \right\rangle \left\langle \left(\Delta \hat{B}\right)^2 \right\rangle \geq \left\{ \left\langle f_1 \left| \left(\hat{A} - \left\langle \hat{A} \right\rangle\right)^\dagger \left(\hat{B} - \left\langle \hat{B} \right\rangle\right) \right| f_1 \right\rangle \right\}^2 \qquad (8.0.31)
$$

where the equality holds only when

$$
\begin{aligned}
g_2 &\propto g_1 \\
g_2 &= c g_1 \\
\left(\hat{B} - \left\langle \hat{B} \right\rangle\right) |f_1\rangle &= c\left(\hat{A} - \left\langle \hat{A} \right\rangle\right) |f_1\rangle \qquad (8.0.32)
\end{aligned}
$$

If we now use $\hat{x}$ and $\hat{p}$ for the operators $\hat{A}$ and $\hat{B}$ then

$$
(\hat{x} - \langle \hat{x} \rangle) |f_1\rangle = c(\hat{p} - \langle \hat{p} \rangle) |f_1\rangle \qquad (8.0.33)
$$

or

$$
(x - \langle \hat{x} \rangle) f_1(x) = c\left(-\imath\hbar \frac{\partial}{\partial x} - \langle \hat{p} \rangle\right) f_1(x) \qquad (8.0.34)
$$

3. Equation (8.0.34) is actually very fundamental. If choose $\langle \hat{x} \rangle = \langle \hat{p} \rangle = 0$ and $c = 1/(\imath\hbar)$, we get the equation

$$
\left(x + \frac{\partial}{\partial x}\right) f_1(x) = \left(\langle \hat{x} \rangle + \frac{\imath \langle \hat{p} \rangle}{\hbar}\right) f_1(x) = 0 \qquad (8.0.35)
$$

You can see that if you pick:

$$f_1(x) = c_1 \exp\left\{ -\frac{(x - \langle \hat{x} \rangle)^2}{4\{\Delta x\}^2} \right\} \exp\left\{ \frac{\imath \langle \hat{p} \rangle x}{\hbar} \right\} \tag{8.0.36}$$

that would satisfy Eq. (8.0.35). (Show this for **homework** Comment on what Δx must be for this to be true. Remember we have picked $\langle \hat{x} \rangle = \langle \hat{p} \rangle = 0$. Eq. (8.0.36) is a gaussian multiplied by a plane wave and is called a *coherent state*. The operator on the left hand side of Eq. (8.0.35), $\left(x + \frac{\partial}{\partial x} \right)$ is called the annihilation operator which we will come across when we do the Harmonic oscillator problem. It *annihilates* the state $f_1(x)$ and the result is zero.

4. *Hence the coherent state is fundamental. It has the minimum uncertainty product. And hence is the most classical-like function in quantum mechanics!! (Note in classical mechanics there is no uncertainty and hence a function that has the minimum value is closest to classical mechanics.) In addition the coherent state is also the eigenstate of the annihilation operator (which we have not discussed yet).*