

UNCERTAINTY PRINCIPLE AND COMPATIBLE OBSERVABLES

B. Zwiebach

October 21, 2013

Contents

1 Uncertainty defined	1
2 The Uncertainty Principle	3
3 The Energy-Time uncertainty	6
4 Lower bounds for ground state energies	9
5 Diagonalization of Operators	11
6 The Spectral Theorem	12
7 Simultaneous Diagonalization of Hermitian Operators	16
8 Complete Set of Commuting Observables	18

1 Uncertainty defined

As we know, observables are associated to Hermitian operators. Given one such operator A we can use it to measure some property of the physical system, as represented by a state Ψ . If the state is in an eigenstate of the operator A , we have no uncertainty in the value of the observable, which coincides with the eigenvalue corresponding to the eigenstate. We only have uncertainty in the value of the observable if the physical state is not an eigenstate of A , but rather a superposition of various eigenstates with different eigenvalues.

We want to define the **uncertainty** $\Delta A(\Psi)$ of the Hermitian operator A on the state Ψ . This uncertainty should vanish if and only if the state is an eigenstate of A . The uncertainty, moreover, should be a real number. In order to define such uncertainty we first recall that the expectation value of A on the state Ψ , assumed to be normalized, is given by

$$\langle A \rangle = \langle \Psi | A | \Psi \rangle = \langle \Psi, A \Psi \rangle. \quad (1.1)$$

The expectation $\langle A \rangle$ is guaranteed to be real since A is Hermitian. We then define the uncertainty as the norm of the vector obtained by acting with $(A - \langle A \rangle I)$ on the physical state (I is the identity operator):

$$\boxed{\Delta A(\Psi) \equiv \left| (A - \langle A \rangle I) \Psi \right|.} \quad (1.2)$$

The uncertainty, so defined is manifestly non-negative. If the uncertainty is zero, the vector inside the norm is zero and therefore:

$$\Delta A(\Psi) = 0 \rightarrow (A - \langle A \rangle I)\Psi = 0 \rightarrow A\Psi = \langle A \rangle \Psi, \quad (1.3)$$

and the last equation confirms that the state is indeed an eigenstate of A (note that $\langle A \rangle$ is a number). You should also note that $\langle A \rangle$ is indeed the eigenvalue, since taking the eigenvalue equation $A\Psi = \lambda\Psi$ and forming the inner product with another Ψ we get

$$\langle \Psi, A\Psi \rangle = \lambda \langle \Psi, \Psi \rangle = \lambda \rightarrow \lambda = \langle A \rangle. \quad (1.4)$$

Alternatively, if the state Ψ is an eigenstate, we now know that the eigenvalue is $\langle A \rangle$ and therefore the state $(A - \langle A \rangle I)\Psi$ vanishes and its norm is zero. We have therefore shown that

The uncertainty $\Delta A(\Psi)$ vanishes if and only if Ψ is an eigenstate of A . (1.5)

To compute the uncertainty one usually squares the expression in (1.2) so that

$$(\Delta A(\Psi))^2 = \langle (A - \langle A \rangle I)\Psi, (A - \langle A \rangle I)\Psi \rangle \quad (1.6)$$

Since the operator A is assumed to be Hermitian and consequently $\langle A \rangle$ is real, we have $(A - \langle A \rangle I)^\dagger = A - \langle A \rangle I$, and therefore we can move the operator on the first entry onto the second one to find

$$(\Delta A(\Psi))^2 = \langle \Psi, (A - \langle A \rangle I)^2 \Psi \rangle. \quad (1.7)$$

While this is a reasonable form, we can simplify it further by expansion

$$(\Delta A(\Psi))^2 = \langle \Psi, (A^2 - 2\langle A \rangle A + \langle A \rangle^2 I)\Psi \rangle. \quad (1.8)$$

The last two term combine and we find

$$(\Delta A(\Psi))^2 = \langle A^2 \rangle - \langle A \rangle^2. \quad (1.9)$$

Since the left-hand side is greater than or equal to zero, this incidentally shows that the expectation value of A^2 is larger than the expectation value of A , squared:

$$\langle A^2 \rangle \geq \langle A \rangle^2. \quad (1.10)$$

An interesting geometrical interpretation of the uncertainty goes as follows. Consider the one-dimensional vector subspace U_Ψ generated by Ψ . Take the state $A\Psi$ and project it to the subspace U_Ψ . The projection, we claim is $\langle A \rangle \Psi$ and the part of $A\Psi$ in the orthogonal subspace U_Ψ^\perp is a vector of norm equal to the uncertainty ΔA . Indeed the orthogonal projector P_{U_Ψ} is

$$P_{U_\Psi} = |\Psi\rangle\langle\Psi|, \quad (1.11)$$

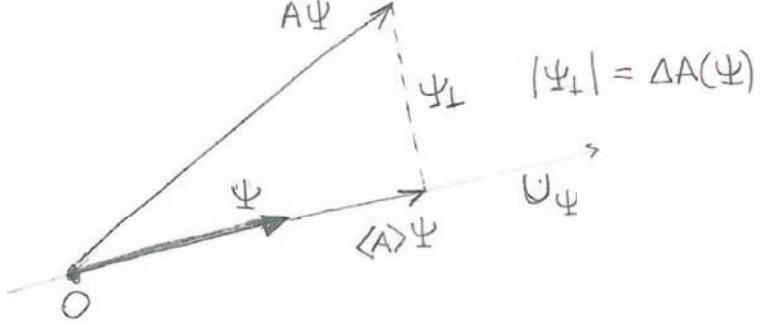


Figure 1: A state Ψ and the one-dimensional subspace U_Ψ generated by it. The projection of $A\Psi$ to U_Ψ is $\langle A \rangle \Psi$. The orthogonal complement Ψ_\perp is a vector whose norm is the uncertainty $\Delta A(\Psi)$.

so that

$$P_{U_\Psi} A|\Psi\rangle = |\Psi\rangle\langle\Psi|A|\Psi\rangle = |\Psi\rangle\langle A|. \quad (1.12)$$

Moreover, the vector $A|\Psi\rangle$ minus its projection must be a vector $|\Psi_\perp\rangle$ orthogonal to $|\Psi\rangle$

$$A|\Psi\rangle - \langle A|\Psi\rangle = |\Psi_\perp\rangle, \quad (1.13)$$

as is easily confirmed by taking the overlap with the bra Ψ . Since the norm of the above left-hand side is the uncertainty, we confirm that $\Delta A = |\Psi_\perp|$, as claimed. These results are illustrated in Figure 1.

2 The Uncertainty Principle

The uncertainty principle is an inequality that is satisfied by the product of the uncertainties of two Hermitian operators that fail to commute. Since the uncertainty of an operator on any given physical state is a number greater than or equal to zero, the product of uncertainties is also a real number greater than or equal to zero. The uncertainty inequality often gives us a lower bound for this product.

When the two operators in question commute, the uncertainty inequality gives no information.

Let us state the uncertainty inequality. Consider two Hermitian operators A and B and a physical state Ψ of the quantum system. Let ΔA and ΔB denote the uncertainties of A and B , respectively, in the state Ψ . Then we have

$$(\Delta A)^2(\Delta B)^2 \geq \left(\langle \Psi | \frac{1}{2i}[A, B] |\Psi \rangle \right)^2. \quad (2.14)$$

The left hand side is a real, non-negative number. For this to be consistent inequality, the right-hand side must also be a real number that is not negative. Since the right-hand side appears squared, the object inside the parenthesis must be real. This can only happen for all Ψ if the operator

$$\frac{1}{2i}[A, B] \quad (2.15)$$

is Hermitian. For this first note that the commutator of two Hermitian operators is *anti*-Hermitian:

$$[A, B]^\dagger = (AB)^\dagger - (BA)^\dagger = B^\dagger A^\dagger - A^\dagger B^\dagger - BA = -[A, B] \quad (2.16)$$

The presence of the i then makes the operator in (2.15) Hermitian. Note that the uncertainty inequality can also be written as

$$\Delta A \Delta B \geq \left| \langle \Psi | \frac{1}{2i} [A, B] | \Psi \rangle \right|. \quad (2.17)$$

where the bars on the right-hand side denote absolute value.

Before we prove the theorem, let's do the canonical example! Substituting \hat{x} for A and \hat{p} for B results in the position-momentum uncertainty relation you have certainly worked with:

$$(\Delta x)^2 (\Delta p)^2 \geq \left(\langle \Psi | \frac{1}{2i} [\hat{x}, \hat{p}] | \Psi \rangle \right)^2. \quad (2.18)$$

Since $[\hat{x}, \hat{p}]/(2i) = \hbar/2$ we get

$$(\Delta x)^2 (\Delta p)^2 \geq \frac{\hbar^2}{4} \rightarrow \Delta x \Delta p \geq \frac{\hbar}{2}. \quad (2.19)$$

We are interested in the proof of the uncertainty inequality for it gives the information that is needed to find the conditions that lead to saturation.

Proof. We define the following two states:

$$\begin{aligned} |f\rangle &\equiv (A - \langle A \rangle I)|\Psi\rangle \\ |g\rangle &\equiv (B - \langle B \rangle I)|\Psi\rangle. \end{aligned} \quad (2.20)$$

Note that by the definition (1.2) of uncertainty,

$$\begin{aligned} \langle f|f\rangle &= (\Delta A)^2, \\ \langle g|g\rangle &= (\Delta B)^2. \end{aligned} \quad (2.21)$$

The Schwarz inequality immediately furnishes us an inequality involving precisely the uncertainties

$$\langle f|f\rangle \langle g|g\rangle \geq |\langle f|g\rangle|^2, \quad (2.22)$$

and therefore we have

$$(\Delta A)^2 (\Delta B)^2 \geq |\langle f|g\rangle|^2 = (\text{Re}\langle f|g\rangle)^2 + (\text{Im}\langle f|g\rangle)^2. \quad (2.23)$$

Writing $\check{A} = (A - \langle A \rangle I)$ and $\check{B} = (B - \langle B \rangle I)$, we now begin to compute the right-hand side:

$$\langle f|g\rangle = \langle \Psi | \check{A} \check{B} | \Psi \rangle = \langle \Psi | (A - \langle A \rangle I)(B - \langle B \rangle I) | \Psi \rangle = \langle \Psi | AB | \Psi \rangle - \langle A \rangle \langle B \rangle, \quad (2.24)$$

and since $|f\rangle$ and $|g\rangle$ go into each other as we exchange A and B ,

$$\langle g|f\rangle = \langle \Psi | \check{A} \check{B} | \Psi \rangle = \langle \Psi | BA | \Psi \rangle - \langle B \rangle \langle A \rangle. \quad (2.25)$$

From the two equations above we find a nice expression for the imaginary part of $\langle f|g \rangle$:

$$\text{Im}\langle f|g \rangle = \frac{1}{2i}(\langle f|g \rangle - \langle g|f \rangle) = \frac{1}{2i}\langle \Psi|[A, B]|\Psi \rangle. \quad (2.26)$$

For the real part the expression is not that simple, so it is best to leave it as the anticommutator of the checked operators:

$$\text{Re}\langle f|g \rangle = \frac{1}{2}(\langle f|g \rangle + \langle g|f \rangle) = \frac{1}{2}\langle \Psi|\{\hat{A}, \hat{B}\}|\Psi \rangle \quad (2.27)$$

Back in (2.23) we get

$$(\Delta A)^2(\Delta B)^2 \geq \left(\langle \Psi|\frac{1}{2i}[A, B]|\Psi \rangle\right)^2 + \left(\langle \Psi|\frac{1}{2}\{\hat{A}, \hat{B}\}|\Psi \rangle\right)^2. \quad (2.28)$$

This can be viewed as the most complete form of the uncertainty inequality. It turns out, however, that the second term on the right hand side is seldom simple enough to be of use, and many times it can be made equal to zero for certain states. At any rate, the term is positive or zero so it can be dropped while preserving the inequality. This is often done, thus giving the celebrated form (2.14) that we have now established.

Now that we have a proven the uncertainty inequality, we can ask: What are the conditions for this inequality to be saturated? If the goal is to minimize uncertainties, under what conditions can we achieve the minimum possible product of uncertainties? As the proof shows, saturation is achieved under two conditions:

1. The Schwarz inequality is saturated. For this we need $|g\rangle = \beta|f\rangle$ where $\beta \in \mathbb{C}$.
2. $\text{Re}(\langle f|g \rangle) = 0$, so that the last term in (2.28) vanishes. This means that $\langle f|g \rangle + \langle g|f \rangle = 0$.

Using $|g\rangle = \beta|f\rangle$ in Condition 2, we get

$$\langle f|g \rangle + \langle g|f \rangle = \beta\langle f|f \rangle + \beta^*\langle f|f \rangle = (\beta + \beta^*)\langle f|f \rangle = 0, \quad (2.29)$$

which requires $\beta + \beta^* = 0$ or that the real part of β vanish. It follows that β must be purely imaginary. So, $\beta = i\lambda$, with λ real, and therefore the uncertainty inequality will be saturated if and only if

$$|g\rangle = i\lambda|f\rangle, \quad \lambda \in \mathbb{R}. \quad (2.30)$$

More explicitly this requires

$$\text{Saturation Condition: } (B - \langle B \rangle I)|\Psi \rangle = i\lambda(A - \langle A \rangle I)|\Psi \rangle. \quad (2.31)$$

This must be viewed as a condition for Ψ , given any two operators A and B . Moreover, note that $\langle A \rangle$ and $\langle B \rangle$ are Ψ dependent. What is λ , physically? Well, the norm of λ is actually fixed by the equation. Taking the norm of both sides we get

$$\Delta B = |\lambda|\Delta A \rightarrow |\lambda| = \frac{\Delta B}{\Delta A}. \quad (2.32)$$

The classic illustration of this saturation condition is worked out for the x, p uncertainty inequality $\Delta x \Delta p \geq \hbar/2$. You will find that gaussian wavefunctions satisfy the saturation condition.

3 The Energy-Time uncertainty

A more subtle form of the uncertainty relation deals with energy and time. The inequality is sometimes stated vaguely in the form $\Delta E \Delta t \gtrsim \hbar$. In here there is no problem in defining ΔE precisely, after all we have the Hamiltonian operator, and its uncertainty ΔH is a perfect candidate for the ‘energy uncertainty’. The problem is time. Time is not an operator in quantum mechanics, it is a parameter, a real number used to describe the way systems change. Unless we define Δt in a precise way we cannot hope for a well-defined uncertainty relation.

We can try a rough, heuristic definition, in order to illustrate the spirit of the inequality. Consider a photon that is detected at some point in space, as a passing oscillatory wave of exact duration T . Even without quantum mechanical considerations we can ask the observer what was the angular frequency ω of the pulse. In order to answer our question the observer will attempt to count the number N of complete oscillations of the waveform that went through. Of course, this number N is given by T divided by the period $2\pi/\omega$ of the wave:

$$N = \frac{\omega T}{2\pi}. \quad (3.33)$$

The observer, however, will typically fail to count full waves, because as the pulse gets started from zero and later on dies off completely, the waveform will cease to follow the sinusoidal pattern. Thus we expect an uncertainty $\Delta N \gtrsim 1$. Given the above relation, this implies an uncertainty $\Delta\omega$ in the value of the angular frequency

$$\Delta\omega T \gtrsim 2\pi. \quad (3.34)$$

This is all still classical, the above identity is something electrical engineers are well aware of. It represents a limit on the ability to ascertain accurately the frequency of a wave that is observed for a limited amount of time. This becomes quantum mechanical if we speak of a single photon, whose energy is $E = \hbar\omega$. Then $\Delta E = \hbar\Delta\omega$, so that multiplying the above inequality by \hbar we get

$$\Delta E T \gtrsim \hbar. \quad (3.35)$$

In this uncertainty inequality T is the duration of the pulse. It is a reasonable relation but the presence of \gtrsim betrays its lack of full precision.

We can find a precise energy/ Q -ness uncertainty inequality by applying the general uncertainty inequality to the Hamiltonian H and another Hermitian operator Q , as did the distinguished Russian physicists L. Mandelstam and Tamm shortly after the formulation of the uncertainty principle. We would then have

$$\Delta H \Delta Q \geq \left| \langle \Psi | \frac{1}{2i} [H, Q] | \Psi \rangle \right|. \quad (3.36)$$

This starting point is interesting because the commutator $[H, Q]$ encodes something very physical about Q . Indeed, let us consider henceforth the case in which the operator Q has no time dependence. It could be, for example some function of \hat{x} and \hat{p} , or for a spin-1/2 particle, the operator $|+\rangle\langle-$. Such

operator Q can easily have time-dependent expectation values, but the time dependence originates from the time dependence of the states, not from the operator Q itself.

To explore the meaning of $[H, Q]$ we begin by computing the time-derivative of the expectation value of Q :

$$\frac{d}{dt}\langle Q \rangle = \frac{d}{dt}\langle \Psi, Q\Psi \rangle = \left\langle \frac{\partial \Psi}{\partial t}, Q\Psi \right\rangle + \left\langle \Psi, Q \frac{\partial \Psi}{\partial t} \right\rangle \quad (3.37)$$

where we did not have to differentiate Q as it is time-independent. At this point we can use the Schrödinger equation to find

$$\begin{aligned} \frac{d}{dt}\langle Q \rangle &= \left\langle \frac{1}{i\hbar} H\Psi, Q\Psi \right\rangle + \left\langle \Psi, Q \frac{1}{i\hbar} H\Psi \right\rangle \\ &= \frac{i}{\hbar} \left(\langle H\Psi, Q\Psi \rangle - \langle \Psi, QH\Psi \rangle \right) \\ &= \frac{i}{\hbar} \langle \Psi, (HQ - QH)\Psi \rangle = \frac{i}{\hbar} \langle \Psi, [H, Q]\Psi \rangle \end{aligned} \quad (3.38)$$

where we used the Hermiticity of the Hamiltonian. We have thus arrived at

$$\boxed{\frac{d}{dt}\langle Q \rangle = \frac{i}{\hbar} \langle [H, Q] \rangle \quad \text{for time-independent } Q.} \quad (3.39)$$

This is a very important result. Each time you see $[H, Q]$ you should think ‘time derivative of $\langle Q \rangle$ ’. In classical mechanics one usually looks for conserved quantities, that is, functions of the dynamical variables that are time independent. In quantum mechanics a conserved operator is one whose expectation value is time independent. An operator Q is conserved if it commutes with the Hamiltonian!

With this result, the inequality (3.36) can be simplified. Indeed, using (3.39) we have

$$\left| \left\langle \frac{1}{2i}[H, Q] \right\rangle \right| = \left| \frac{1}{2i} \frac{\hbar}{i} \frac{d\langle Q \rangle}{dt} \right| = \frac{\hbar}{2} \left| \frac{d\langle Q \rangle}{dt} \right| \quad (3.40)$$

and therefore

$$\boxed{\Delta H \Delta Q \geq \frac{\hbar}{2} \left| \frac{d\langle Q \rangle}{dt} \right|, \quad \text{for time-independent } Q.} \quad (3.41)$$

This is a perfectly precise uncertainty inequality. The terms in it suggest a definition of a time Δt_Q

$$\Delta t_Q \equiv \frac{\Delta Q}{\left| \frac{d\langle Q \rangle}{dt} \right|}. \quad (3.42)$$

This quantity has units of time. It is the time it would take $\langle Q \rangle$ to change by ΔQ if both ΔQ and the velocity $\frac{d\langle Q \rangle}{dt}$ were time-independent. Since they are not necessarily so, we can view Δt_Q as the time for ‘appreciable’ change in $\langle Q \rangle$. This is certainly so when $\langle Q \rangle$ and ΔQ are roughly of the same size. In terms of Δt_Q the uncertainty inequality reads

$$\Delta H \Delta t_Q \geq \frac{\hbar}{2}. \quad (3.43)$$

This is still a precise inequality, given that Δt_Q has a concrete definition in (3.42).

As you will consider in the homework, (3.41) can be used to derive an inequality for time Δt_{\perp} that it takes for a system to become orthogonal to itself. If we call the initial state $\Psi(0)$, we call Δt_{\perp} the smallest time for which $\langle \Psi(0), \Psi(\Delta t_{\perp}) \rangle = 0$. You will be able to show that

$$\Delta H \Delta t_{\perp} \geq \frac{\hbar}{4}. \quad (3.44)$$

The speed in which a state can turn orthogonal depends on the energy uncertainty, and in quantum computation it plays a role in limiting the maximum possible speed of a computer for a fixed finite energy.

The uncertainty relation involves ΔH . It is natural to ask if this quantity is time dependent. As we show now, it is not, if the Hamiltonian is a time-independent operator. Indeed, if H is time independent, we can use H and H^2 for Q in (3.39) so that

$$\begin{aligned} \frac{d}{dt} \langle H \rangle &= \frac{i}{\hbar} \langle [H, H] \rangle = 0, \\ \frac{d}{dt} \langle H^2 \rangle &= \frac{i}{\hbar} \langle [H, H^2] \rangle = 0. \end{aligned} \quad (3.45)$$

It then follows that

$$\frac{d}{dt} (\Delta H)^2 = \frac{d}{dt} (\langle H^2 \rangle - \langle H \rangle^2) = 0. \quad (3.46)$$

showing that ΔH is a constant. So we have shown that

If H is time independent, the uncertainty ΔH is constant in time. (3.47)

The concept of conservation of energy uncertainty can be used to understand some aspects of atomic decays. Consider, for illustration the hyperfine transition in the hydrogen atom. Due to the existence of proton spin and the electron spin, the ground state of hydrogen is fourfold degenerate, corresponding to the four possible combinations of spins (up-up, up-down, down-up, down-down). The magnetic interaction between the spins actually breaks this degeneracy and produces the so-called “hyperfine” splitting. This is a very tiny split: $\delta E = 5.88 \times 10^{-6}$ ev (compare with about 13.6 ev for the ground state energy). For a hyperfine atomic transition, the emitted photon carries the energy difference δE resulting in a wavelength of 21.1cm and a frequency $\nu = 1420.405751786(30)$ MHz. The eleven significant digits of this frequency attest to the sharpness of the emission line. The issue of uncertainty arises because the excited state of the hyperfine splitting has a lifetime τ_H for decay to the ground state and emission of a photon. This lifetime is extremely long, in fact $\tau_H \sim 11$ million years ($= 3.4 \times 10^{14}$ sec, recalling that a year is about $\pi \times 10^7$ sec, accurate to better than 1%). This lifetime can be viewed as the time that takes some observable of the electron-proton system to change significantly (its total spin angular momentum, perhaps) so by the uncertainty principle it must be related to some energy uncertainty $\Delta E \sim \hbar/\tau_H \simeq 2 \times 10^{-30}$ ev. of the original excited state of the

hydrogen atom. Once the decay takes place the atom goes to the fully stable ground state, without any possible energy uncertainty. By the conservation of energy uncertainty, the photon must carry the uncertainty ΔE . But $\Delta E/\delta E \sim 3 \times 10^{-25}$, an absolutely infinitesimal effect on the photon. There is no broadening of the 21 cm line! That's one reason it is so useful in astronomy. For decays with much shorter lifetimes there can be an observable broadening of an emission line due to the energy-time uncertainty principle.

4 Lower bounds for ground state energies

You may recall that the variational principle could be used to find *upper* bounds on ground state energies. The uncertainty principle can be used to find *lower* bounds for the ground state energy of certain systems. Use below the uncertainty principle in the form $\Delta x \Delta p \geq \hbar/2$ to find rigorous lower bounds for the ground state energy of one-dimensional Hamiltonians. This is best illustrated by an example.

Consider a particle in a one-dimensional quartic potential considered earlier

$$H = \frac{p^2}{2m} + \alpha x^4, \quad (4.48)$$

where $\alpha > 0$ is a constant with units of energy over length to the fourth power. Our goal is to find a *lower bound* for the ground state energy $\langle H \rangle_{gs}$. Taking the ground state expectation value of the Hamiltonian we have

$$\langle H \rangle_{gs} = \frac{\langle p^2 \rangle_{gs}}{2m} + \alpha \langle x^4 \rangle_{gs}, \quad (4.49)$$

Recalling that

$$(\Delta p)^2 = \langle p^2 \rangle - \langle p \rangle^2, \quad (4.50)$$

we see that

$$\langle p^2 \rangle \geq (\Delta p)^2, \quad (4.51)$$

for any state of the system. We should note however, that for the ground state (or any bound state) $\langle p \rangle = 0$ so that in fact

$$\langle p^2 \rangle_{gs} = (\Delta p)_{gs}^2, \quad (4.52)$$

From the inequality $\langle A^2 \rangle \geq \langle A \rangle^2$ we have

$$\langle x^4 \rangle \geq \langle x^2 \rangle^2. \quad (4.53)$$

Moreover, just like for momentum above, $(\Delta x)^2 = \langle x^2 \rangle - \langle x \rangle^2$ leads to

$$\langle x^2 \rangle \geq (\Delta x)^2, \quad (4.54)$$

so that

$$\langle x^4 \rangle \geq (\Delta x)^4, \quad (4.55)$$

for the expectation value on arbitrary states. Therefore

$$\langle H \rangle_{gs} = \frac{\langle p^2 \rangle_{gs}}{2m} + \alpha \langle x^4 \rangle_{gs} \geq \frac{(\Delta p_{gs})^2}{2m} + \alpha (\Delta x_{gs})^4 \quad (4.56)$$

From the uncertainty principle

$$\Delta x_{gs} \Delta p_{gs} \geq \frac{\hbar}{2} \rightarrow \Delta p_{gs} \geq \frac{\hbar}{2\Delta x_{gs}}. \quad (4.57)$$

Back to the value of $\langle H \rangle_{gs}$ we get

$$\langle H \rangle_{gs} \geq \frac{\hbar^2}{8m(\Delta x_{gs})^2} + \alpha (\Delta x_{gs})^4. \quad (4.58)$$

The quantity to the right of the inequality is a function of Δx_{gs} . This function has been plotted in Figure 2.

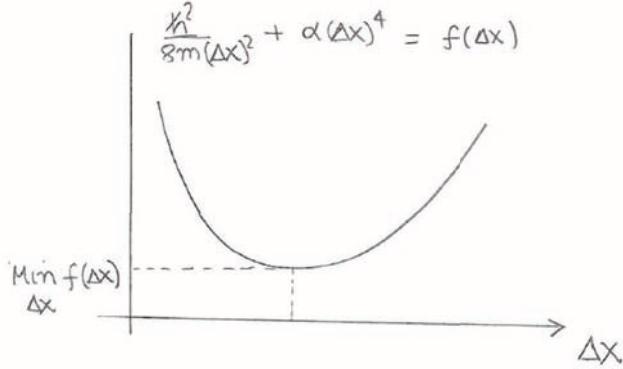


Figure 2: We have that $\langle H_{gs} \rangle \geq f(\Delta x_{gs})$ but we don't know the value of Δx_{gs} . As a result, we can only be certain that $\langle H_{gs} \rangle$ is greater than or equal to the *lowest* value the function $f(\Delta x_{gs})$ can take.

If we knew the value of Δx_{gs} we would immediately know that $\langle H \rangle_{gs}$ is bigger than the value taken by the right-hand side. This would be quite nice, since we want the highest possible lower bound. Since we don't know the value of Δx_{gs} , however, the only thing we can be sure of is that $\langle H \rangle_{gs}$ is bigger than the *lowest* value that can be taken by the expression to the right of the inequality as we vary Δx_{gs} :

$$\langle H \rangle_{gs} \geq \text{Min}_{\Delta x} \left(\frac{\hbar^2}{8m(\Delta x)^2} + \alpha (\Delta x)^4 \right). \quad (4.59)$$

The minimization problem is straightforward. In fact

$$f(x) = \frac{A}{x^2} + Bx^4 \text{ is minimized for } x^2 = 2^{-1/3} \left(\frac{A}{B} \right)^{1/3} \text{ yielding } f = 2^{1/3} \frac{3}{2} (A^2 B)^{1/3}. \quad (4.60)$$

Applied to (4.59) we obtain

$$\langle H \rangle_{gs} \geq 2^{1/3} \frac{3}{8} \left(\frac{\hbar^2 \sqrt{\alpha}}{m} \right)^{2/3} \simeq 0.4724 \left(\frac{\hbar^2 \sqrt{\alpha}}{m} \right)^{2/3}. \quad (4.61)$$

This is the final lower bound for the ground state energy. It is actually not too bad, for the ground state instead of the prefactor 0.4724, we have 0.668.

5 Diagonalization of Operators

When we have operators we wish to understand, it can be useful to find a basis on the vector space for which the operators are represented by matrices that take a simple form. Diagonal matrices are matrices where all non diagonal entries vanish. If we can find a set of basis vectors for which the matrix representing an operator is diagonal we say that the operator is **diagonalizable**.

If an operator T is diagonal in some basis (u_1, \dots, u_n) of the vector space V , its matrix takes the form $\text{diag}(\lambda_1, \dots, \lambda_n)$, with constants λ_i , and we have

$$Tu_1 = \lambda_1 u_1, \dots, Tu_n = \lambda_n u_n. \quad (5.62)$$

The basis vectors are recognized as eigenvectors with eigenvalues given by the diagonal elements. It follows that *a matrix is diagonalizable if and only if it possesses a set of eigenvectors that span the vector space*. Recall that all operators T on complex vector spaces have at least one eigenvalue and thus at least one eigenvector. But not even in complex vector spaces all operators have enough eigenvectors to span the space. Those operators cannot be diagonalized. The simplest example of such operator is provided by the two-by-two matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (5.63)$$

The only eigenvalue of this matrix is $\lambda = 0$ and the associated eigenvector is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Since a two-dimensional vector space cannot be spanned with one eigenvector, this matrix cannot be diagonalized. Having seen that the question of diagonalization of an operator is ultimately a question about its eigenvectors, we want to emphasize that the question can be formulated without referring to any basis. Bases, of course are useful, to express concretely

Suppose we have a vector space V and we have chosen a basis (v_1, \dots, v_n) such that a linear operator has a matrix representation $T_{ij}(\{v\})$ that is not diagonal. As we learned before, if we change basis to a new one (u_1, \dots, u_n) using a linear operator A such that

$$u_k = A v_k, \quad (5.64)$$

the matrix representation $T_{ij}(\{u\})$ of the operator in the new basis takes the form

$$T(\{u\}) = A^{-1} T(\{v\}) A \quad \text{or} \quad T_{ij}(\{u\}) = (A^{-1})_{ik} T_{kp}(\{v\}) A_{pj}, \quad (5.65)$$

where the matrix A_{ij} is the representation of A in the original v -basis. The operator T is diagonalizable if there is an operator A such that $T_{ij}(\{u\})$ is diagonal.

There are two pictures of the diagonalization: One can consider the operator T and state that its matrix representation is diagonal when referred to the u basis obtained by acting with A on the original v basis. Alternatively, we can view the result as the existence of a related operator $A^{-1}TA$ that is diagonal in the *original* v basis. Indeed, $Tu_i = \lambda_i u_i$ (i not summed) implies that $TA v_i = \lambda_i Av_i$

and acting with A^{-1} that $(A^{-1}TA)v_i = \lambda_i v_i$, which confirms that $A^{-1}TA$ is represented by a diagonal matrix in the original v basis. Both viewpoints are valuable.

It is useful to note that the columns of the matrix A are in fact the eigenvectors of $T(\{v\})$. We see this as follows. Since the eigenvectors are the u_k we have

$$u_k = Av_k \quad \rightarrow \quad u_k = \sum_i A_{ik} v_i . \quad (5.66)$$

Using the original basis means v_i is represented by a column vector of zeroes with a single unit entry at the i -th position. We thus find

$$u_k = \begin{pmatrix} A_{1k} \\ \vdots \\ A_{nk} \end{pmatrix} . \quad (5.67)$$

confirming that the k -th column of A is the k -th eigenvector of T .

While not all operators on complex vector spaces can be diagonalized, the situation is much improved for Hermitian operators. Recall that T is Hermitian if $T = T^\dagger$. Hermitian operators can be diagonalized, and so can unitary operators. But even more is true: the operators take diagonal form in an orthonormal basis!

An operator M is said to be **unitarily** diagonalizable if there is an *orthonormal* basis in which its matrix representation is a diagonal matrix. That basis, therefore, is an orthonormal basis of eigenvectors. Starting with an arbitrary orthonormal basis (e_1, \dots, e_n) where the matrix representation of M is $M(\{e\})$, a unitary transformation of this basis produces the orthonormal basis in which the operator takes diagonal form. More explicitly, there is a unitary matrix U ($U^\dagger = U^{-1}$) and a diagonal matrix D_M such that

$$U^\dagger M(\{e\}) U = D_M . \quad (5.68)$$

6 The Spectral Theorem

While we could prove, as most textbooks do, that Hermitian operators are unitarily diagonalizable, this result holds for a more general class of operators, called normal operators. The proof is not harder than the one for hermitian operators. An operator M is said to be **normal** if it commutes with its adjoint:

$$M \text{ is normal : } [M^\dagger, M] = 0 . \quad (6.69)$$

Hermitian operators are clearly normal. So are anti-hermitian operators ($M^\dagger = -M$ is antihermitian). Unitary operators U are normal because both $U^\dagger U$ and UU^\dagger are equal to the identity matrix and thus U and U^\dagger commute.

Exercise. If an operator M is normal show that so is $V^\dagger M V$ where V is a unitary operator.

Lemma: Let w be an eigenvector of the normal operator M : $Mw = \lambda w$. Then w is also an eigenvector of M^\dagger with complex conjugate eigenvalue:

$$M^\dagger w = \lambda^* w. \quad (6.70)$$

Proof: Define $u = (M^\dagger - \lambda^* I)w$. The result holds if u is the zero vector. To show this we compute the norm-squared of u :

$$|u|^2 = \langle u, u \rangle = \langle (M^\dagger - \lambda^* I)w, (M^\dagger - \lambda^* I)w \rangle \quad (6.71)$$

Using the adjoint property to move the operator in the first entry to the second entry:

$$|u|^2 = \langle w, (M - \lambda I)(M^\dagger - \lambda^* I)w \rangle \quad (6.72)$$

Since M and M^\dagger commute, so do the two factors in parenthesis and therefore

$$|u|^2 = \langle w, (M^\dagger - \lambda^* I)(M - \lambda I)w \rangle = 0 \quad (6.73)$$

since $(M - \lambda I)$ kills w . It follows that $u = 0$ and therefore (6.70) holds. \square

We can now state our main theorem, called the *spectral theorem*. It states that a matrix is unitarily diagonalizable if and only if it is normal. More to the point,

Spectral Theorem: Let M be an operator in a complex vector space. The vector space has a orthonormal basis comprised of eigenvectors of M if and only if M is normal.

(6.74)

Proof. It is easy to show that unitarily diagonalizable implies normality. Indeed, from (5.68) and dropping the reference to the e -basis,

$$M = UD_M U^\dagger \text{ and therefore } M^\dagger = U D_M^\dagger U^\dagger.$$

We then get

$$M^\dagger M = U D_M^\dagger D_M U^\dagger \text{ and } M M^\dagger = U D_M D_M^\dagger U^\dagger.$$

so that

$$[M^\dagger, M] = U(D_M^\dagger D_M - D_M D_M^\dagger)U^\dagger = 0,$$

because any two diagonal matrices commute.

Now let us prove that M provides a basis of orthonormal eigenvectors. The proof is by induction. The result is clearly true for $\dim V = 1$. We assume that it holds for $(n-1)$ -dimensional vector spaces and consider the case of n -dimensional V . Let M be an $n \times n$ matrix referred to the orthonormal basis $(|1\rangle, \dots, |n\rangle)$ of V so that $M_{ij} = \langle i|M|j \rangle$. We know there is at least one eigenvalue λ_1 with a non-zero eigenvector $|x_1\rangle$ of unit norm:

$$M|x_1\rangle = \lambda_1|x_1\rangle \text{ and } M^\dagger|x_1\rangle = \lambda_1^*|x_1\rangle, \quad (6.75)$$

in view of the Lemma. There is, we claim, a unitary matrix U_1 such that

$$|x_1\rangle = U_1|1\rangle \rightarrow U_1^\dagger|x_1\rangle = |1\rangle. \quad (6.76)$$

U_1 is not unique and can be constructed as follows: extend $|x_1\rangle$ to an orthonormal basis $|x_1\rangle, \dots, |x_N\rangle$ using Gram-Schmidt. Then write $U_1 = \sum_i |x_i\rangle\langle i|$. Define now

$$M_1 \equiv U_1^\dagger M U_1. \quad (6.77)$$

M_1 is also normal and $M_1|1\rangle = U_1^\dagger M U_1|1\rangle = U_1^\dagger M|x_1\rangle = \lambda_1 U_1^\dagger|x_1\rangle = \lambda_1|1\rangle$, so that

$$M_1|1\rangle = \lambda_1|1\rangle. \quad (6.78)$$

Let us now examine the explicit form of the matrix M_1 :

$$\langle j|M_1|1\rangle = \lambda_1\langle j|1\rangle = \lambda_1\delta_{i,j}, \quad (6.79)$$

which says that the first column of M_1 has zeroes in all entries except the first. Moreover

$$\langle 1|M_1|j\rangle = (\langle j|M_1^\dagger|1\rangle)^* = (\lambda_1^*\langle j|1\rangle)^* = \lambda_1\langle 1|j\rangle = \lambda_1\delta_{i,j}, \quad (6.80)$$

where we used $M_1^\dagger|1\rangle = \lambda_1^*|1\rangle$ which follows from the normality of M_1 . It follows from the two last equations that M_1 , in the original basis, takes the form

$$M_1 = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & M' & \\ 0 & & & \end{pmatrix}.$$

Since M_1 is normal, one can see that M' is a normal $(n-1)$ -by- $(n-1)$ matrix. By the induction hypothesis M' can be unitarily diagonalized so that $U'^\dagger M' U'$ is diagonal for some $(n-1)$ -by- $(n-1)$ unitary matrix U' . The matrix U' can be extended to an n -by- n unitary matrix \hat{U} as follows

$$\hat{U} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & U' & \\ 0 & & & \end{pmatrix}. \quad (6.81)$$

It follows that $\hat{U}^\dagger M_1 \hat{U} = \hat{U}^\dagger U_1^\dagger M U_1 \hat{U} = (U_1 \hat{U})^\dagger M (U_1 \hat{U})$ is diagonal, proving the desired result. \square .

Of course this theorem implies that Hermitian and unitary operators are unitarily diagonalizable. In other words the eigenvectors form an orthonormal basis. This is true whether or not there are degeneracies in the spectrum. The proof does not require discussion of this as a special case. If an eigenvalue of M is degenerate and appears k times, then there are k orthonormal eigenvectors associated with the corresponding k -dimensional M -invariant subspace of the vector space.

- - - - -

We conclude this section with a description of the general situation that we may encounter when diagonalizing a normal operator T . In general, we expect degeneracies in the eigenvalues so that each eigenvalue λ_k is repeated $d_k \geq 1$ times. An eigenvalue λ_k is degenerate if $d_k > 1$. It follows that V has T -invariant subspaces of different dimensionalities. Let U_k denote the T -invariant subspace of dimension $d_k \geq 1$ spanned by eigenvectors with eigenvalue λ_k :

$$U_k \equiv \{v \in V \mid T v = \lambda_k v\}, \quad \dim U_k = d_k. \quad (6.82)$$

By the spectral theorem U_k has a basis comprised by d_k orthonormal eigenvectors $(u_1^{(k)}, \dots, u_{d_k}^{(k)})$. Note that while the addition of eigenvectors with different eigenvalues does not give eigenvectors, in the subspace U_k all vectors are eigenvectors with the same eigenvalue, and that's why addition makes sense, U_k as defined is a vector space, and adding eigenvectors in U_k gives eigenvectors. The full space V is decomposed as the direct sum of the invariant subspaces of T :

$$V = U_1 \oplus U_2 \oplus \dots \oplus U_m, \quad \dim V = \sum_{i=1}^m d_i, \quad m \geq 1. \quad (6.83)$$

All U_i subspaces are guaranteed to be orthogonal to each other. In fact the full list of eigenvectors is a list of orthonormal vectors that form a basis for V is conveniently ordered as follows:

$$(u_1^{(1)}, \dots, u_{d_1}^{(1)}, \dots, u_1^{(m)}, \dots, u_{d_m}^{(m)}). \quad (6.84)$$

The matrix T is manifestly diagonal in this basis because each vector above is an eigenvector of T and is orthogonal to all others. The matrix representation of T reads

$$T = \text{diag} \left(\underbrace{\lambda_1, \dots, \lambda_1}_{d_1 \text{ times}}, \dots, \underbrace{\lambda_m, \dots, \lambda_m}_{d_m \text{ times}} \right) \quad (6.85)$$

This is clear because the first d_1 vectors in the list are in U_1 , the second d_2 vectors are in U_2 , and so on and so forth until the last d_m vectors are in U_m .

If we had no degeneracies in the spectrum the basis (6.84) (with $d_i = 1$ for all i) would be rather unique if we require the matrix representation of T to be unchanged. Each vector could be multiplied by a phase. On the other hand, with degeneracies that the list (6.84) can be changed considerably without changing the matrix representation of T . Let V_k be a unitary operator on U_k , for each $k = 1, \dots, m$. We claim that the following basis of eigenvectors leads to the same matrix T :

$$(V_1 u_1^{(1)}, \dots, V_1 u_{d_1}^{(1)}, \dots, V_m u_1^{(m)}, \dots, V_m u_{d_m}^{(m)}). \quad (6.86)$$

Indeed, this is still a collection of eigenvectors of T with each of them orthogonal to the rest. Moreover, the first d_1 vectors are in U_1 , the second d_2 vectors are in U_2 and so on and so forth. More explicitly, for example, within U_k

$$\langle V_k u_i^{(k)}, T(V_k u_j^{(k)}) \rangle = \lambda_k \langle V_k u_i^{(k)}, V_k u_j^{(k)} \rangle = \lambda_k \langle u_i^{(k)}, u_j^{(k)} \rangle = \lambda_k \delta_{ij} \quad (6.87)$$

showing that in the U_k subspace the matrix for T is still diagonal with all entries equal to λ_k .

7 Simultaneous Diagonalization of Hermitian Operators

We say that two operators S and T in a vector space V operators can be **simultaneously diagonalized** if there is some basis of V in which both the matrix representation of S and the matrix representation of T are diagonal. It then follows that each vector in this basis is an eigenvector of S and an eigenvector of T .

A necessary condition for simultaneous diagonalization is that the operators S and T commute. Indeed, if they can be simultaneously diagonalized there is a basis where both are diagonal and they manifestly commute. If the operators don't commute, this is a basis-independent statement and therefore a simultaneous diagonal presentation cannot exist. Since arbitrary linear operators S and T on a complex vector space cannot be diagonalized, the vanishing of $[S, T]$ does not guarantee simultaneous diagonalization. But if the operators are Hermitian it does, as we show now.

Theorem. If S and T are commuting Hermitian operators they can be simultaneously diagonalized.

Proof. The main complication is that degeneracies in the spectrum require an some discussion. Either both operators have degeneracies or one has no degeneracies. Without loss of generality we can assume that there are two cases to consider

- (i) There is no degeneracy in the spectrum of T or,
- (ii) Both T and S have degeneracies in their spectrum.

Consider case (i) first. Since T is non-degenerate there is a basis (u_1, \dots, u_n) of eigenvectors of T with different eigenvalues

$$Tu_i = \lambda_i u_i, \quad i \text{ not summed}, \quad \lambda_i \neq \lambda_j \text{ for } i \neq j. \quad (7.88)$$

We now want to understand what kind of vector is Su_i . For this we act with T on it

$$T(Su_i) = S(Tu_i) = S(\lambda_i u_i) = \lambda_i (Su_i), \quad (7.89)$$

It follows that Su_i is also an eigenvector of T with eigenvalue λ_i , thus it must equal u_i , up to scale,

$$Su_i = \omega_i u_i, \quad (7.90)$$

showing that u_i is also an eigenvector of S , this time with eigenvalue ω_i . Thus any eigenvector of T is also an eigenvector of S , showing that these operators are simultaneously diagonalizable.

Now consider case (ii). Since T has degeneracies, as explained in the previous section, we have a decomposition of V in T -invariant subspaces U_k spanned by eigenvectors:

$$\begin{aligned} U_k &\equiv \{u \mid Tu = \lambda_k u\}, \quad \dim U_k = d_k \quad V = U_1 \oplus \dots \oplus U_m, \\ \text{orthonormal basis for } V : \quad &(u_1^{(1)}, \dots, u_{d_1}^{(1)}, \dots, u_1^{(m)}, \dots, u_{d_m}^{(m)}). \\ T &= \text{diag} \left(\underbrace{\lambda_1, \dots, \lambda_1}_{d_1 \text{ times}}, \dots, \underbrace{\lambda_m, \dots, \lambda_m}_{d_m \text{ times}} \right) \text{ in this basis.} \end{aligned} \quad (7.91)$$

We also explained that the alternative orthonormal basis of V

$$(V_1 u_1^{(1)}, \dots V_1 u_{d_1}^{(1)}, \dots, V_m u_1^{(m)}, \dots, V_m u_{d_m}^{(m)}). \quad (7.92)$$

leads to the same matrix for T when each V_k is a unitary operator on U_k .

We now claim that the U_k are also S -invariant subspaces! To show this let $u \in U_k$ and examine the vector Su . We have

$$T(Su) = S(Tu) = \lambda_k Su \rightarrow Su \in U_k. \quad (7.93)$$

We use the subspaces U_k and the basis (7.91) to organize the matrix representation of S in blocks. It follows that this matrix must have *block-diagonal* form since each subspace is S -invariant and orthogonal to all other subspaces. We cannot guarantee, however, that S is diagonal within each square block because $Su_i^{(k)} \in U_k$ but we have no reason to believe that $Su_i^{(k)}$ points along $u_i^{(k)}$.

Since S restricted to each S -invariant subspace U_k is hermitian we can find an orthonormal basis of U_k in which the matrix S is diagonal. This new basis is unitarily related to the original basis $(u_1^{(k)}, \dots, u_{d_k}^{(k)})$ and thus takes the form $(V_k u_1^{(k)}, \dots, V_k u_{d_k}^{(k)})$ with V_k a unitary operator in U_k . Note that the eigenvalues of S in this block need not be degenerate. Doing this for each block, we find a basis of the form (7.92) in which S is diagonal. But T is still diagonal in this new basis, so both S and T have been simultaneously diagonalized. \square

Remarks:

1. Note that the above proof gives an algorithmic way to produce the common list of eigenvectors. One diagonalizes one of the matrices and constructs the second matrix in the basis of eigenvectors of the first. These second matrix is block diagonal, where the blocks are organized by the degeneracies in the spectrum of the first matrix. One must then diagonalize within the blocks and is guaranteed that the new basis that works for the second matrix also works for the first.
2. If we had to simultaneously diagonalize three different commuting Hermitian operators S_1, S_2 and S_3 , all of which have degenerate spectra, we would proceed as follows. We diagonalize S_1 and fix a basis in which S_1 is diagonal. In this basis we must find that S_2 and S_3 have exactly the same block structure. The corresponding block matrices are simply the matrix representations of S_2 and S_3 in each of the invariant spaces U_k appearing in the diagonalization of S_1 . Since S_2 and S_3 commute, their restrictions to U_k commute. These restrictions can be diagonalized simultaneously, as guaranteed by our theorem which works for two matrices. The new basis in U_k that makes the restriction of S_2 and S_3 diagonal, will not disturb the diagonal form of S_1 in this block. This is repeated for each block, until we get a common basis of eigenvectors.
3. An inductive algorithm is clear. If we know how to simultaneously diagonalize n commuting Hermitian operators we can diagonalize $n + 1$ of them, call them S_1, \dots, S_{n+1} , as follows. We diagonalize S_1 and then consider the remaining n operators in the basis that makes S_1 diagonal.

We are guaranteed a common block structure for the n operators. The problem becomes one of simultaneous diagonalization of n commuting Hermitian block matrices, which is assumed known by the induction argument.

Corollary. If $\{S_1, \dots, S_n\}$ is a set of mutually commuting Hermitian operators they can all be simultaneously diagonalized.

8 Complete Set of Commuting Observables

We have discussed the problem of finding eigenvectors and eigenvalues of a Hermitian operator S . This hermitian operator is thought as a quantum mechanical observable. The eigenvectors of S are physical states of the system in which the observable S can be measured without uncertainty. The result of the measurement is the eigenvalue associated with the eigenvector.

If the Hermitian operator S has a non-degenerate spectrum, all eigenvalues are different and we have a rather nice situation in which each eigenvector can be uniquely distinguished by labeling it with the corresponding eigenvalue of S . The physical quantity associated with the observable can be used to distinguish the various eigenstates. Moreover, these eigenstates provide an orthonormal basis for the full vector space. In this case the operator S provides a “complete set of commuting observables” or a CSCO, in short. The set here has just one observable, the operator S .

The situation is more nontrivial if the Hermitian operator S exhibits degeneracies in its spectrum. This means that V has an S -invariant subspace of dimension $d > 1$, spanned by orthonormal eigenvectors (u_1, \dots, u_d) all of which have S eigenvalue λ . This time, the eigenvalue of S does not allow us to distinguish or to label uniquely the basis eigenstates of the invariant subspace. Physically this is a deficient situation, as we have explicitly different states – the various u_i ’s – that we can’t tell apart by the measurement of S alone. This time S does not provide a CSCO. Labeling eigenstates by the S eigenvalue does not suffice to distinguish them.

We are thus physically motivated to find another Hermitian operator T that is compatible with S . Two Hermitian operators are said to be **compatible observables** if they commute, since then we can find a basis of V comprised by simultaneous eigenvectors of the operators. These states can be labeled by two observables, namely, the two eigenvalues. If we are lucky, the basis eigenstates in each of the S -invariant subspaces of dimension higher than one can be organized into T eigenstates of different eigenvalues. In this case T breaks the spectral degeneracy of S and using T eigenvalues as well as S eigenvalues we can label uniquely a basis of orthonormal states of V . In this case we say that S and T form a CSCO.

We have now given enough motivation for a definition of a complete set of commuting observables. Consider a set of commuting observables, namely, a set $\{S_1, \dots, S_k\}$ of Hermitian operators acting on a complex vector space V that represents the physical state-space of some quantum system. By the theorem in the previous section, we can find an orthonormal basis of vectors in V such that each vector

is an eigenstate of every operator in the set. Assume that each eigenstate in the basis is labeled by the eigenvalues of the S_i operators. The set $\{S_1, \dots, S_k\}$ is said to be a **complete set of commuting observables** if no two states have the same labels.

It is a physically motivated assumption that for any physical quantum system there is a complete set of commuting observables, for otherwise there is no physical way to distinguish the various states that span the vector space. So in any physical problem we are urged to find such complete set, and we must include operators in such set until all degeneracies are broken. A CSCO need not be unique. Once we have a complete set of commuting observables, adding another observable causes no harm, although it is not necessary. Also, if (S_1, S_2) form a CSCO, so will $(S_1 + S_2, S_1 - S_2)$. Ideally, we want the smallest set of operators.

The first operator that is usually included in a CSCO is the Hamiltonian H . For bound state problems in one dimension, energy eigenstates are non-degenerate and thus the energy can be used to label uniquely the H -eigenstates. A simple example is the infinite square well. Another example is the one-dimensional harmonic oscillator. In such cases H forms the CSCO. If we have, however, a two-dimensional isotropic harmonic oscillator in the (x, y) plane, the Hamiltonian has degeneracies. At the first excited level we can have the first excited state of the x harmonic oscillator or, at the same energy, the first excited state of the y harmonic oscillator. We thus need another observable that can be used to distinguish these states. There are several options, as you will discuss in the homework.

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 10: Solving the Time-Independent Schrödinger Equation

B. Zwiebach
March 14, 2016

Contents

1 Stationary States	1
2 Solving for Energy Eigenstates	3
3 Free particle on a circle.	6

1 Stationary States

Consider the Schrödinger equation for the wavefunction $\Psi(x, t)$ with the assumption that the potential energy V is time independent:

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi(x, t) = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right) \Psi(x, t), \quad (1.1)$$

where we displayed the form of the Hamiltonian operator \hat{H} with the time independent potential $V(x)$. Stationary states are a very useful class of solutions of this differential equation. The signature property of a stationary state is that the position and the time dependence of the wavefunction factorize. Namely,

$$\Psi(x, t) = g(t)\psi(x), \quad (1.2)$$

for some functions g and ψ . For such a *separable* solution to exist we need the potential to be time independent, as we will see below. The solution $\Psi(x, t)$ is time-dependent but it is called stationary because of a property of observables. The expectation value of observables with no explicit time dependence in arbitrary states has time dependence. On a stationary state they do not have time dependence, as we will demonstrate.

Let us use the ansatz (1.2) for Ψ in the Schrödinger equation. We then find

$$\left(i\hbar \frac{dg(t)}{dt} \right) \psi(x) = g(t) \hat{H}\psi(x), \quad (1.3)$$

because $g(t)$ can be moved across \hat{H} . We can then divide this equation by $\Psi(x, t) = g(t)\psi(x)$, giving

$$i\hbar \frac{1}{g(t)} \frac{dg(t)}{dt} = \frac{1}{\psi(x)} \hat{H}\psi(x). \quad (1.4)$$

The left side is a function of only t , while the right side is a function of only x (a time dependent potential would have spoiled this). The only way the two sides can equal each other for all values of t and x is for both sides to be equal to a *constant E* with units of energy because \hat{H} has units of energy. We therefore get two separate equations. The first reads

$$i\hbar \frac{dg}{dt} = Eg. \quad (1.5)$$

This is solved by

$$g(t) = e^{-iEt/\hbar}, \quad (1.6)$$

and the most general solution is simply a constant times the above right-hand side. From the x -dependent side of the equality we get

$$\hat{H}\psi(x) = E\psi(x). \quad (1.7)$$

This equation is an eigenvalue equation for the Hermitian operator \hat{H} . We showed that the eigenvalues of Hermitian operators must be real, thus the constant E **must be real**. The equation above is called the **time-independent Schrödinger equation**. More explicitly it reads

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right) \psi(x) = E\psi(x), \quad (1.8)$$

Note that this equation does not determine the overall normalization of ψ . Therefore we can write the full solution without loss of generality using the $g(t)$ given above:

$$\text{Stationary state: } \Psi(x, t) = e^{-iEt/\hbar} \psi(x), \quad \text{with } E \in \mathbb{R} \text{ and } \hat{H}\psi = E\psi. \quad (1.9)$$

Note that not only is $\psi(x)$ an eigenstate of the Hamiltonian operator \hat{H} , the full stationary state is also an \hat{H} eigenstate

$$\hat{H}\Psi(x, t) = E\Psi(x, t), \quad (1.10)$$

since the time dependent function in Ψ cancels out.

We have noted that the energy E must be real. If it was not we would also have trouble normalizing the stationary state consistently. The normalization condition for Ψ , if E is not real, would give

$$\begin{aligned} 1 &= \int dx \Psi^*(x, t)\Psi(x, t) = \int dx e^{iE^*t/\hbar} e^{-iEt/\hbar} \psi^*(x)\psi(x) \\ &= e^{i(E^*-E)t/\hbar} \int dx \psi^*(x)\psi(x) = e^{2\text{Im}(E)t/\hbar} \int dx \psi^*(x)\psi(x). \end{aligned} \quad (1.11)$$

The final expression has a time dependence due to the exponential. On the other hand the normalization condition states that this expression must be equal to one. It follows that the exponent must be zero, i.e., E is real. Given this, we also see that the normalization condition yields

$$\int_{-\infty}^{\infty} dx \psi^*(x)\psi(x) = 1. \quad (1.12)$$

How do we interpret the eigenvalue E ? Using (1.10) we see that the expectation value of \hat{H} on the state Ψ is indeed the energy

$$\langle \hat{H} \rangle_{\Psi} = \int dx \Psi^*(x, t)\hat{H}\Psi(x, t) = \int dx \Psi^*(x, t)E\Psi(x, t) = E \int dx \Psi^*(x, t)\Psi(x, t) = E, \quad (1.13)$$

Since the stationary state is an eigenstate of \hat{H} , the uncertainty $\Delta\hat{H}$ of the Hamiltonian in a stationary state is zero.

There are two important observations on stationary states:

- (1) The expectation value of any time-independent operator \hat{Q} on a stationary state Ψ is time-independent:

$$\begin{aligned}\langle Q \rangle_{\Psi(x,t)} &= \int dx \Psi^*(x,t) \hat{Q} \Psi(x,t) = \int dx e^{iEt/\hbar} \psi^*(x) \hat{Q} e^{-iEt/\hbar} \psi(x) \\ &= \int dx e^{iEt/\hbar} e^{-iEt/\hbar} \psi^*(x) \hat{Q} \psi(x) = \int dx \psi^*(x) \hat{Q} \psi(x) = \langle Q \rangle_{\psi(x)},\end{aligned}\tag{1.14}$$

since the last expectation value is manifestly time independent.

- (2) The superposition of stationary states with different energies not stationary. This is clear because a stationary state requires a factorized solution of the Schrödinger equation: if we add two factorized solutions with different energies they will have different time dependence and the total state cannot be factorized. We now show that that a time-independent observable \hat{Q} may have a time-dependent expectation values in such a state. Consider a superposition

$$\Psi(x,t) = c_1 e^{-iE_1 t/\hbar} \psi_1(x) + c_2 e^{-iE_2 t/\hbar} \psi_2(x),\tag{1.15}$$

where ψ_1 and ψ_2 are \hat{H} eigenstates with energies E_1 and E_2 , respectively. Consider a Hermitian operator \hat{Q} . With the system in state (1.15), its expectation value is

$$\begin{aligned}\langle Q \rangle_{\Psi} &= \int_{-\infty}^{\infty} dx \Psi^*(x,t) \hat{Q} \Psi(x,t) \\ &= \int_{-\infty}^{\infty} dx (c_1^* e^{iE_1 t/\hbar} \psi_1^*(x) + c_2^* e^{iE_2 t/\hbar} \psi_2^*(x)) (c_1 e^{-iE_1 t/\hbar} \hat{Q} \psi_1(x) + c_2 e^{-iE_2 t/\hbar} \hat{Q} \psi_2(x)) \\ &= \int_{-\infty}^{\infty} dx \left(|c_1|^2 \psi_1^* \hat{Q} \psi_1 + |c_2|^2 \psi_2^* \hat{Q} \psi_2 + c_1^* c_2 e^{i(E_1 - E_2)t/\hbar} \psi_1^* \hat{Q} \psi_2 + c_2^* c_1 e^{-i(E_1 - E_2)t/\hbar} \psi_2^* \hat{Q} \psi_1 \right)\end{aligned}\tag{1.16}$$

We now see the possible time dependence arising from the cross terms. The first two terms are simple time-independent expectation values. Using the hermitically of \hat{Q} in the last term we then get

$$\begin{aligned}\langle Q \rangle_{\Psi} &= |c_1|^2 \langle Q \rangle_{\psi_1} + |c_2|^2 \langle Q \rangle_{\psi_2} \\ &\quad + c_1^* c_2 e^{i(E_1 - E_2)t/\hbar} \int_{-\infty}^{\infty} dx \psi_1^* \hat{Q} \psi_2 + c_1 c_2^* e^{-i(E_1 - E_2)t/\hbar} \int_{-\infty}^{\infty} dx \psi_1 (\hat{Q} \psi_2)^*\end{aligned}\tag{1.17}$$

The last two terms are complex conjugates of each other and therefore

$$\langle Q \rangle_{\Psi} = |c_1|^2 \langle Q \rangle_{\psi_1} + |c_2|^2 \langle Q \rangle_{\psi_2} + 2 \operatorname{Re} \left[c_1^* c_2 e^{i(E_1 - E_2)t/\hbar} \int_{-\infty}^{\infty} dx \psi_1^* \hat{Q} \psi_2 \right].\tag{1.18}$$

We see that this expectation value is time-dependent if $E_1 \neq E_2$ and $(\psi_1, Q \psi_2)$ is nonzero. The full expectation value $\langle Q \rangle_{\Psi}$ is real, as it must be for any Hermitian operator.

2 Solving for Energy Eigenstates

We will now study solutions to the time-independent Schrödinger equation

$$\hat{H} \psi(x) = E \psi(x).\tag{2.19}$$

For a given Hamiltonian \hat{H} we are interested in finding the eigenstates ψ and the eigenvalues E , which happen to be the corresponding energies. Perhaps the most interesting feature of the above equation is that generally the value of E cannot be arbitrary. Just like finite size matrices have a set of eigenvalues, the above, time-independent Schrödinger equation may have a discrete set of possible energies. A continuous set of possible energies is also allowed and sometimes important. There are indeed many solutions for any given potential. Assuming for convenience that the eigenstates and their energies can be counted we write

$$\begin{aligned} \psi_1(x), & \quad E_1 \\ \psi_2(x), & \quad E_2 \\ \vdots & \quad \vdots \end{aligned} \tag{2.20}$$

Our earlier discussion of Hermitian operators applies here. The energy eigenstates can be organized to form a *complete set of orthonormal functions*:

$$\int \psi_i^*(x) \psi_j(x) = \delta_{ij}. \tag{2.21}$$

Consider the time-independent Schrödinger equation written as

$$\frac{d^2\psi}{dx^2} = -\frac{2m}{\hbar^2}(E - V(x))\psi. \tag{2.22}$$

The solutions $\psi(x)$ depend on the properties of the potential $V(x)$. It is hard to make general statements about the wavefunction unless we restrict the types of potentials. We will certainly consider continuous potentials. We also consider potentials that are not continuous but are piece-wise continuous, that is, they have a number of discontinuities. Our potentials can easily fail to be bounded. We allow delta functions in one-dimensional potentials but do not consider powers or derivatives of delta functions. We allow for potentials that become plus infinity beyond certain points. These points represent hard walls.

We want to understand general properties of ψ and the behavior of ψ at points where the potential $V(x)$ may have discontinuities or other singularities. We claim: **we must have a continuous wavefunction**. If ψ is discontinuous then ψ' contains delta-functions and ψ'' in the above left-hand side contains derivatives of delta functions. This would require the right-hand side to have derivatives of delta functions, and those would have to appear in the potential. Since we have declared that our potentials contain no derivatives of delta functions we must indeed have a continuous ψ .

Consider now four possibilities concerning the potential:

- (1) $V(x)$ is continuous. In this case the continuity of $\psi(x)$ and (2.22) imply ψ'' is also continuous. This requires ψ' continuous.
- (2) $V(x)$ has finite discontinuities. In this case ψ'' has finite discontinuities: it includes the product of a continuous ψ against a discontinuous V . But then ψ' must be continuous, with non-continuous derivative.
- (3) $V(x)$ contains delta functions. In this case ψ'' also contains delta functions: it is proportional to the product of a continuous ψ and a delta function in V . Thus ψ' has finite discontinuities.

(4) $V(x)$ contains a hard wall. A potential that is finite immediately to the left of $x = a$ and becomes infinite for $x > a$ is said to have a hard wall at $x = a$. In such a case, the wavefunction will vanish for $x \geq a$. The slope ψ' will be finite as $x \rightarrow a$ from the left, and will vanish for $x > a$. Thus ψ' is discontinuous at the wall.

In the first two cases ψ' is continuous, and in the second two it can have a finite discontinuity. In conclusion

Both ψ and ψ' are continuous unless the potential has delta functions or hard walls in which cases ψ' may have finite discontinuities. (2.23)

Let us give an slightly different argument for the continuity of ψ and $\frac{d\psi}{dx}$ in the case of a potential with a finite discontinuity, such as the step shown in Fig. 1.

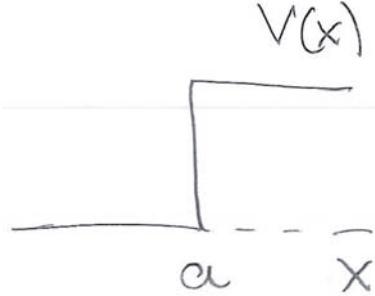


Figure 1: A potential $V(x)$ with a finite discontinuity at $x = a$.

Integrate both sides of (2.22) $a - \epsilon$ to $a + \epsilon$, and then take $\epsilon \rightarrow 0$. We find

$$\int_{a-\epsilon}^{a+\epsilon} dx \frac{d}{dx} \left(\frac{d\psi}{dx} \right) = -\frac{2m}{\hbar^2} \int_{a-\epsilon}^{a+\epsilon} dx (E - V(x))\psi(x). \quad (2.24)$$

The left-hand side integrand is a total derivative so we have

$$\frac{d\psi}{dx} \Big|_{a+\epsilon} - \frac{d\psi}{dx} \Big|_{a-\epsilon} = \frac{2m}{\hbar^2} \int_{a-\epsilon}^{a+\epsilon} dx (V(x) - E)\psi(x). \quad (2.25)$$

By definition, the discontinuity in the derivative of ψ at $x = a$ is the limit as $\epsilon \rightarrow 0$ of the left-hand side:

$$\Delta_a \left(\frac{d\psi}{dx} \right) \equiv \lim_{\epsilon \rightarrow 0} \left(\frac{d\psi}{dx} \Big|_{a+\epsilon} - \frac{d\psi}{dx} \Big|_{a-\epsilon} \right). \quad (2.26)$$

Back in (2.25) we then have

$$\Delta_a \left(\frac{d\psi}{dx} \right) = \lim_{\epsilon \rightarrow 0} \frac{2m}{\hbar^2} \int_{a-\epsilon}^{a+\epsilon} dx (V(x) - E)\psi(x). \quad (2.27)$$

The potential V is discontinuous but not infinite around $x = a$, nor is ψ infinite around $x = a$ and, of course, E is assumed finite. As the integral range becomes vanishingly small about $x = a$ the integrand remains finite and the integral goes to zero. We thus have

$$\Delta_a \left(\frac{d\psi}{dx} \right) = 0. \quad (2.28)$$

There is no discontinuity in $\frac{d\psi}{dx}$. This gives us one of our boundary conditions.

To learn about the continuity of ψ we reconsider the first integral of the differential equation. The integration that led to (2.25) now applied to the range from $x_0 < a$ to x yields

$$\frac{d\psi(x)}{dx} = \left. \frac{d\psi}{dx} \right|_{x_0} - \frac{2m}{\hbar} \int_{x_0}^x (E - V(x')) dx'. \quad (2.29)$$

Note that the integral on the right is a bounded function of x . We now integrate again from $a - \epsilon$ to $a + \epsilon$. Since the first term on the right-hand side is a constant we find

$$\psi(a + \epsilon) - \psi(a - \epsilon) = 2\epsilon \left. \frac{d\psi}{dx} \right|_{x_0} - \frac{2m}{\hbar} \int_{a-\epsilon}^{a+\epsilon} dx \int_{x_0}^x (E - V(x')). \quad (2.30)$$

Taking the $\epsilon \rightarrow 0$ limit, the first term on the right-hand side clearly vanishes and the second term goes to zero because $\int_{x_0}^x dx' (E - V(x'))$ is a bounded function of x . As a result we have

$$\Delta_a \psi = 0, \quad (2.31)$$

showing that the wavefunction is continuous at $x = a$. This is our second boundary condition.

3 Free particle on a circle.

Consider now the problem of a particle confined to a circle of circumference L . The coordinate along the circle is called x and we can view the circle as the interval $x \in [0, L]$ with the endpoints identified. It is perhaps clearer mathematically to think of the circle as the full real line x with the identification

$$x \sim x + L, \quad (3.1)$$

which means that two points whose coordinates are related in this way are to be considered **the same point**. It follows that we have the periodicity condition

$$\psi(x + L) = \psi(x). \quad (3.2)$$

From this it follows that not only ψ is periodic but all of its derivatives are also periodic.

The particle is assumed to be free and therefore $V(x) = 0$. The time-independent Schrödinger equation is then

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} = E \psi(x). \quad (3.3)$$

Before we solve this, let us show that any solution must have $E \geq 0$. For this multiply the above equation by $\psi^*(x)$ and integrate over the circle $x \in [0, L]$. Since ψ is normalized we get

$$-\frac{\hbar^2}{2m} \int_0^L \psi^*(x) \frac{d^2\psi}{dx^2} dx = E \int \psi^*(x) \psi(x) dx = E. \quad (3.4)$$

The integrand on the left hand side can be rewritten as

$$-\frac{\hbar^2}{2m} \int_0^L \left[\frac{d}{dx} \left(\psi^* \frac{d\psi}{dx} \right) - \frac{d\psi^*}{dx} \frac{d\psi}{dx} \right] dx = E. \quad (3.5)$$

and the total derivative can be integrated

$$-\frac{\hbar^2}{2m} \left[\left(\psi^* \frac{d\psi}{dx} \right) \Big|_{x=L} - \left(\psi^* \frac{d\psi}{dx} \right) \Big|_{x=0} \right] + \frac{\hbar^2}{2m} \int_0^L \left| \frac{d\psi}{dx} \right|^2 dx = E. \quad (3.6)$$

Since $\psi(x)$ and its derivatives are periodic, the contributions from $x = L$ and $x = 0$ cancel out and we are left with

$$E = \frac{\hbar^2}{2m} \int_0^L \left| \frac{d\psi}{dx} \right|^2 dx \geq 0, \quad (3.7)$$

which establishes our claim. We also see that $E = 0$ requires ψ constant (and nonzero!).

Having shown that all solutions must have $E \geq 0$ let us go back to the Schrödinger equation, which can be rewritten as

$$\frac{d^2\psi}{dx^2} = -\frac{2mE}{\hbar^2} \psi. \quad (3.8)$$

We can then define k via

$$k^2 \equiv \frac{2mE}{\hbar^2} \geq 0. \quad (3.9)$$

Since $E \geq 0$, the constant k is real. Note that this definition is very natural, since it makes

$$E = \frac{\hbar^2 k^2}{2m}, \quad (3.10)$$

which means that, as usual, $p = \hbar k$. Using k^2 the differential equation becomes the familiar

$$\frac{d^2\psi}{dx^2} = -k^2 \psi. \quad (3.11)$$

We could write the general solution in terms of sines and cosines of kx , but let's use complex exponentials:

$$\psi(x) \sim e^{ikx}. \quad (3.12)$$

This solves the differential equation and, moreover, it is a momentum eigenstate. The periodicity condition (3.2) requires

$$e^{ik(x+L)} = e^{ikx} \rightarrow e^{ikL} = 1 \rightarrow kL = 2\pi n, \quad n \in \mathbb{Z}. \quad (3.13)$$

We see that momentum is quantized because the wavenumber is quantized! The wavenumber has discrete possible values

$$k_n \equiv \frac{2\pi n}{L}, \quad n \in \mathbb{Z}. \quad (3.14)$$

All integers positive and negative are allowed and are in fact necessary because they all correspond to *different* values of the momentum $p_n = \hbar k_n$. The solutions to the Schrödinger equation can then be indexed by the integer n :

$$\psi_n(x) = N e^{ik_n x}, \quad (3.15)$$

where N is a real normalization constant. Its value is determined from

$$1 = \int_0^L \psi_n^*(x) \psi_n(x) dx = \int_0^L N^2 dx = N^2 L \rightarrow N = \frac{1}{\sqrt{L}}, \quad (3.16)$$

so we have

$$\psi_n(x) = \frac{1}{\sqrt{L}} e^{ik_n x} = \frac{1}{\sqrt{L}} e^{\frac{2\pi i n x}{L}}. \quad (3.17)$$

The associated energies are

$$E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2 4\pi^2 n^2}{2m L^2} = \frac{2\pi^2 \hbar^2 n^2}{m L^2}. \quad (3.18)$$

There are infinitely many energy eigenstates. We have degenerate states because E_n is just a function of $|n|$ and thus the same for n and $-n$. Indeed ψ_n and ψ_{-n} both have energy E_n . The only nondegenerate eigenstate is $\psi_0 = \frac{1}{\sqrt{L}}$, which is a constant wavefunction with zero energy.

Whenever we find degenerate energy eigenstates we must wonder what makes those states different, given that they have the same energy. To answer this one must find an observable that takes different values on the states. Happily, in our case we know the answer. Our degenerate states can be distinguished by their momentum: ψ_n has momentum $2\pi n \frac{\hbar}{L}$ and ψ_{-n} has momentum $(-2\pi n \frac{\hbar}{L})$.

Given two degenerate energy eigenstates, any linear combination of these states is an eigenstate with the same energy. Indeed if

$$\hat{H}\psi_1 = E\psi_1, \quad \hat{H}\psi_2 = E\psi_2, \quad (3.19)$$

then

$$\hat{H}(a\psi_1 + b\psi_2) = a\hat{H}\psi_1 + b\hat{H}\psi_2 = aE\psi_1 + bE\psi_2 = E(a\psi_1 + b\psi_2). \quad (3.20)$$

We can therefore form two linear combinations of the degenerate eigenstates ψ_n and ψ_{-n} to obtain another description of the energy eigenstates:

$$\begin{aligned} \psi_n + \psi_{-n} &\sim \cos(k_n x), \\ \psi_n - \psi_{-n} &\sim \sin(k_n x). \end{aligned} \quad (3.21)$$

While these are real energy eigenstates, they are not momentum eigenstates. Only our exponentials are simultaneous eigenstates of both \hat{H} and \hat{p} .

The energy eigenstates ψ_n are automatically orthonormal since they are \hat{p} eigenstates with no degeneracies (and as you recall eigenstates of a hermitian operator with different eigenvalues are automatically orthogonal) :

$$\int_0^L \psi_n^*(x) \psi_m(x) dx = \frac{1}{L} \int_0^L e^{\frac{2\pi i (m-n)x}{L}} dx = \delta_{mn}. \quad (3.22)$$

They are also complete: we can then construct a general wavefunction as a superposition that is in fact a Fourier series. For any $\Psi(x, 0)$ that satisfies the periodicity condition, we can write

$$\Psi(x, 0) = \sum_{n \in \mathbb{Z}} a_n \psi_n(x), \quad (3.23)$$

where, as you should check, the coefficients a_n are determined by the integrals

$$a_n = \int_0^L dx \psi_n^*(x) \Psi(x, 0). \quad (3.24)$$

The initial state $\Psi(x, 0)$ is then easily evolved in time:

$$\Psi(x, t) = \sum_{n \in \mathbb{Z}} a_n \psi_n(x) e^{-\frac{i E_n t}{\hbar}}. \quad (3.25)$$

Andrew Turner transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare

<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

LINEAR ALGEBRA: VECTOR SPACES AND OPERATORS

B. Zwiebach

October 21, 2013

Contents

1	Vector spaces and dimensionality	1
2	Linear operators and matrices	5
3	Eigenvalues and eigenvectors	11
4	Inner products	14
5	Orthonormal basis and orthogonal projectors	18
6	Linear functionals and adjoint operators	20
7	Hermitian and Unitary operators	24

1 Vector spaces and dimensionality

In quantum mechanics the state of a physical system is a *vector* in a *complex* vector space. Observables are linear operators, in fact, Hermitian operators acting on this complex vector space. The purpose of this chapter is to learn the basics of vector spaces, the structures that can be built on those spaces, and the operators that act on them.

Complex vector spaces are somewhat different from the more familiar real vector spaces. I would say they have more powerful properties. In order to understand more generally complex vector spaces it is useful to compare them often to their real dimensional friends. We will follow here the discussion of the book *Linear algebra done right*, by Sheldon Axler.

In a vector space one has vectors and numbers. We can add vectors to get vectors and we can multiply vectors by numbers to get vectors. If the numbers we use are real, we have a real vector space. If the numbers we use are complex, we have a complex vector space. More generally, the numbers we use belong to what is called in mathematics a ‘field’ and denoted by the letter \mathbb{F} . We will discuss just two cases, $\mathbb{F} = \mathbb{R}$, meaning that the numbers are real, and $\mathbb{F} = \mathbb{C}$, meaning that the numbers are complex.

The definition of a vector space is the same for \mathbb{F} being \mathbb{R} or \mathbb{C} . A vector space V is a set of vectors with an operation of **addition** (+) that assigns an element $u + v \in V$ to each $u, v \in V$. This means that V is closed under addition. There is also a **scalar multiplication** by elements of \mathbb{F} , with $av \in V$

for any $a \in \mathbb{F}$ and $v \in V$. This means the space V is closed under multiplication by numbers. These operations must satisfy the following additional properties:

1. $u + v = v + u \in V$ for all $u, v \in V$ (addition is commutative).
2. $u + (v + w) = (u + v) + w$ and $(ab)u = a(bu)$ for any $u, v, w \in V$ and $a, b \in \mathbb{F}$ (associativity).
3. There is a vector $0 \in V$ such that $0 + u = u$ for all $u \in V$ (additive identity).
4. For each $v \in V$ there is a $u \in V$ such that $v + u = 0$ (additive inverse).
5. The element $1 \in \mathbb{F}$ satisfies $1v = v$ for all $v \in V$ (multiplicative identity).
6. $a(u + v) = au + av$ and $(a + b)v = av + bv$ for every $u, v \in V$ and $a, b \in \mathbb{F}$ (distributive property).

This definition is very efficient. Several familiar properties follow from it by short proofs (which we will not give, but are not complicated and you may try to produce):

- The additive identity is unique: any vector $0'$ that acts like 0 is actually equal to 0 .
- $0v = 0$, for any $v \in V$, where the first zero is a number and the second one is a vector. This means that the number zero acts as expected when multiplying a vector.
- $a0 = 0$, for any $a \in \mathbb{F}$. Here both zeroes are vectors. This means that the zero vector multiplied by any number is still the zero vector.
- The additive inverse of any vector $v \in V$ is unique. It is denoted by $-v$ and in fact $-v = (-1)v$.

We must emphasize that while the numbers, in \mathbb{F} are sometimes real or complex, we never speak of the vectors themselves as real or complex. A vector multiplied by a complex number is not said to be a complex vector, for example! The vectors in a real vector space are not themselves real, nor are the vectors in a complex vector space complex. We have the following examples of vector spaces:

1. The set of N -component vectors

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}, \quad a_i \in \mathbb{R}, \quad i = 1, 2, \dots, N. \quad (1.1)$$

form a real vector space.

2. The set of $M \times N$ matrices with complex entries

$$\begin{pmatrix} a_{11} & \dots & a_{1N} \\ a_{21} & \dots & a_{2N} \\ \vdots & \vdots & \vdots \\ a_{M1} & \dots & a_{MN} \end{pmatrix}, \quad a_{ij} \in \mathbb{C}, \quad (1.2)$$

is a complex vector space. In here multiplication by a constant multiplies each entry of the matrix by the constant.

3. We can have matrices with complex entries that naturally form a real vector space. The space of two-by-two *hermitian* matrices define a *real* vector space. They do not form a complex vector space since multiplication of a hermitian matrix by a complex number ruins the hermiticity.
4. The set $\mathcal{P}(\mathbb{F})$ of polynomials $p(z)$. Here the variable $z \in \mathbb{F}$ and $p(z) \in \mathbb{F}$. Each polynomial $p(z)$ has coefficients a_0, a_1, \dots, a_n also in \mathbb{F} :

$$p(z) = a_0 + a_1 z + a_2 z^2 + \dots + a_n z^n. \quad (1.3)$$

By definition, the integer n is finite but it can take any nonnegative value. Addition of polynomials works as expected and multiplication by a constant is also the obvious multiplication. The space $\mathcal{P}(\mathbb{F})$ of all polynomials so defined form a vector space over \mathbb{F} .

5. The set \mathbb{F}^∞ of infinite sequences (x_1, x_2, \dots) of elements $x_i \in \mathbb{F}$. Here

$$\begin{aligned} (x_1, x_2, \dots) + (y_1, y_2, \dots) &= (x_1 + y_1, x_2 + y_2, \dots) \\ a(x_1, x_2, \dots) &= (ax_1, ax_2, \dots) \quad a \in \mathbb{F}. \end{aligned} \quad (1.4)$$

This is a vector space over \mathbb{F} .

6. The set of complex functions on an interval $x \in [0, L]$, form a vector space over \mathbb{C} .

To better understand a vector space one can try to figure out its possible subspaces. A **subspace** of a vector space V is a subset of V that is also a vector space. To verify that a subset U of V is a subspace you must check that U contains the vector 0, and that U is closed under addition and scalar multiplication.

Sometimes a vector space V can be described clearly in terms of collection U_1, U_2, \dots, U_m of subspaces of V . We say that the space V is the **direct sum** of the subspaces U_1, U_2, \dots, U_m and we write

$$V = U_1 \oplus U_2 \oplus \dots \oplus U_m \quad (1.5)$$

if any vector in V can be written *uniquely* as the sum $u_1 + u_2 + \dots + u_m$, where $u_i \in U_i$. To check uniqueness one can, alternatively, verify that the only way to write 0 as a sum $u_1 + u_2 + \dots + u_m$ with $u_i \in U_i$ is by taking all u_i 's equal to zero. For the case of two subspaces $V = U \oplus W$, it suffices to prove that any vector can be written as $u + w$ with $u \in U$ and $w \in W$ and that $U \cap W = 0$.

Given a vector space we can produce lists of vectors. A **list** (v_1, v_2, \dots, v_n) of vectors in V contains, by definition, a finite number of vectors. The number of vectors in the list is the length of the list. The **span** of a list of vectors (v_1, v_2, \dots, v_n) in V , denoted as $\text{span}(v_1, v_2, \dots, v_n)$, is the set of all linear combinations of these vectors

$$a_1 v_1 + a_2 v_2 + \dots + a_n v_n, \quad a_i \in \mathbb{F} \quad (1.6)$$

A vector space V is spanned by a list (v_1, v_2, \dots, v_n) if $V = \text{span}(v_1, v_2, \dots, v_n)$.

Now comes a very natural definition: A vector space V is said to be **finite dimensional** if it is spanned by some list of vectors in V . If V is not finite dimensional, it is **infinite dimensional**. In such case, no list of vectors from V can span V .

Let us show that the vector space of all polynomials $p(z)$ considered in Example 4 is an infinite dimensional vector space. Indeed, consider any list of polynomials. In this list there is a polynomial of maximum degree (recall the list is finite). Thus polynomials of higher degree are not in the span of the list. Since no list can span the space, it is infinite dimensional.

For example 1, consider the list of vectors (e_1, e_2, \dots, e_N) with

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots \quad e_N = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \quad (1.7)$$

This list spans the space (the vector displayed is $a_1e_1 + a_2e_2 + \dots + a_Ne_N$). This vector space is finite dimensional.

A list of vectors (v_1, v_2, \dots, v_n) , with $v_i \in V$ is said to be **linearly independent** if the equation

$$a_1v_1 + a_2v_2 + \dots + a_nv_n = 0, \quad (1.8)$$

only has the solution $a_1 = a_2 = \dots = a_n = 0$. One can show that the length of any linearly independent list is shorter or equal to the length of any spanning list. This is reasonable, because spanning lists can be arbitrarily long (adding vectors to a spanning list gives still a spanning list), but a linearly independent list cannot be enlarged beyond a certain point.

Finally, we get to the concept of a basis for a vector space. A **basis** of V is a list of vectors in V that both spans V and it is linearly independent. Mathematicians easily prove that any finite dimensional vector space has a basis. Moreover, all bases of a finite dimensional vector space have the same length. The **dimension** of a finite-dimensional vector space is given by the length of any list of basis vectors. One can also show that for a finite dimensional vector space a list of vectors of length $\dim V$ is a basis if it is linearly independent list or if it is a spanning list.

For example 1 we see that the list (e_1, e_2, \dots, e_N) in (1.7) is not only a spanning list but a linearly independent list (prove it!). Thus the dimensionality of this space is N .

For example 3, recall that the most general hermitian two-by-two matrix takes the form

$$\begin{pmatrix} a_0 + a_3 & a_1 - ia_2 \\ a_1 + ia_2 & a_0 - a_3 \end{pmatrix}, \quad a_0, a_1, a_2, a_3 \in \mathbb{R}. \quad (1.9)$$

Now consider the following list of four ‘vectors’ $(\mathbf{1}, \sigma_1, \sigma_2, \sigma_3)$. All entries in this list are hermitian matrices, so this is a list of vectors in the space. Moreover they span the space since the most general hermitian matrix, as shown above, is simply $a_0\mathbf{1} + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3$. The list is linearly independent

as $a_0\mathbf{1} + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 = 0$ implies that

$$\begin{pmatrix} a_0 + a_3 & a_1 - ia_2 \\ a_1 + ia_2 & a_0 - a_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad (1.10)$$

and you can quickly see that this implies a_0, a_1, a_2 , and a_3 are zero. So the list is a basis and the space in question is a four-dimensional real vector space.

Exercise. Explain why the vector space in example 2 has dimension $M \cdot N$.

It seems pretty obvious that the vector space in example 5 is infinite dimensional, but it actually takes a bit of work to prove it.

2 Linear operators and matrices

A linear map refers in general to a certain kind of function from one vector space V to another vector space W . When the linear map takes the vector space V to itself, we call the linear map a linear operator. We will focus our attention on those operators. Let us then define a linear operator.

A **linear operator** T on a vector space V is a function that takes V to V with the properties:

1. $T(u + v) = Tu + Tv$, for all $u, v \in V$.
2. $T(au) = aTu$, for all $a \in \mathbb{F}$ and $u \in V$.

We call $\mathcal{L}(V)$ the set of all linear operators that act on V . This can be a very interesting set, as we will see below. Let us consider a few examples of linear operators.

1. Let V denote the space of real polynomials $p(x)$ of a real variable x with real coefficients. Here are two linear operators:
 - Let T denote differentiation: $Tp = p'$. This operator is linear because $(p_1 + p_2)' = p'_1 + p'_2$ and $(ap)' = ap'$.
 - Let S denote multiplication by x : $Sp = xp$. S is also a linear operator.
2. In the space \mathbb{F}^∞ of infinite sequences define the left-shift operator L by

$$L(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots). \quad (2.11)$$

We lose the first entry, but that is perfectly consistent with linearity. We also have the right-shift operator R that acts as follows:

$$R(x_1, x_2, \dots) = (0, x_1, x_2, \dots). \quad (2.12)$$

Note that the first entry in the result is zero. It could not be any other number because the zero element (a sequence of all zeroes) should be mapped to itself (by linearity).

3. For any V , the zero map 0 such that $0v = 0$. This map is linear and maps all elements of V to the zero element.
4. For any V , the identity map I for which $Iv = v$ for all $v \in V$. This map leaves all vectors invariant.

Since operators on V can be added and can also be multiplied by numbers, the set $\mathcal{L}(V)$ introduced above is itself a vector space (the vectors being the operators!). Indeed for any two operators $T, S \in \mathcal{L}(V)$ we have the natural definition

$$\begin{aligned} (S + T)v &= Sv + Tv, \\ (aS)v &= a(Sv). \end{aligned} \tag{2.13}$$

The additive identity in the vector space $\mathcal{L}(V)$ is the zero map of example 3.

In this vector space there is a surprising new structure: the vectors (the operators!) can be multiplied. There is a multiplication of linear operators that gives a linear operator. We just let one operator act first and the second later. So given $S, T \in \mathcal{L}(V)$ we define the operator ST as

$$(ST)v \equiv S(Tv) \tag{2.14}$$

You should convince yourself that ST is a linear operator. This product structure in the space of linear operators is associative: $S(TU) = (ST)U$, for S, T, U , linear operators. Moreover it has an identity element: the identity map of example 4. Most crucially this multiplication is, in general, *noncommutative*. We can check this using the two operators T and S of example 1 acting on the polynomial $p = x^n$. Since T differentiates and S multiplies by x we get

$$(TS)x^n = T(Sx^n) = T(x^{n+1}) = (n+1)x^n, \quad \text{while} \quad (ST)x^n = S(Tx^n) = S(nx^{n-1}) = nx^n. \tag{2.15}$$

We can quantify this failure of commutativity by writing the difference

$$(TS - ST)x^n = (n+1)x^n - nx^n = x^n = I x^n \tag{2.16}$$

where we inserted the identity operator at the last step. Since this relation is true for any x^n , it would also hold acting on any polynomial, namely on any element of the vector space. So we write

$$[T, S] = I. \tag{2.17}$$

where we introduced the commutator $[\cdot, \cdot]$ of two operators X, Y , defined as $[X, Y] \equiv XY - YX$.

The most basic features of an operator are captured by two simple concepts: its null space and its range. Given some linear operator T on V it is of interest to consider those elements of V that are mapped to the zero element. The **null space** (or kernel) of $T \in \mathcal{L}(V)$ is the subset of vectors in V that are mapped to zero by T :

$$\text{null } T = \{v \in V; Tv = 0\}. \tag{2.18}$$

Actually $\text{null } T$ is a *subspace* of V (The only nontrivial part of this proof is to show that $T(0) = 0$. This follows from $T(0) = T(0 + 0) = T(0) + T(0)$ and then adding to both sides of this equation the additive inverse to $T(0)$).

A linear operator $T : V \rightarrow V$ is said to be **injective** if $Tu = Tv$, with $u, v \in V$, implies $u = v$. An injective map is called a *one-to-one* map, because not two different elements can be mapped to the same one. In fact, physicist Sean Carroll has suggested that a better name would be *two-to-two* as injectivity really means that two different elements are mapped by T to two different elements! We leave for you as an exercise to prove the following important characterization of injective maps:

Exercise. Show that T is injective if and only if $\text{null } T = \{0\}$.

Given a linear operator T on V it is also of interest to consider the elements of V of the form Tv . The linear operator may not produce by its action all of the elements of V . We define the **range** of T as the image of V under the map T :

$$\text{range } T = \{Tv; v \in V\}. \quad (2.19)$$

Actually $\text{range } T$ is a *subspace* of V (can you prove it?). The linear operator T is said to be **surjective** if $\text{range } T = V$. That is, if the image of V under T is the complete V .

Since both the null space and the range of a linear operator $T : V \rightarrow V$ are subspaces of V , one can assign a dimension to them, and the following theorem is nontrivial:

$$\dim V = \dim(\text{null } T) + \dim(\text{range } T). \quad (2.20)$$

Example. Describe the null space and range of the operator

$$T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (2.21)$$

Let us now consider invertible linear operators. A linear operator $T \in \mathcal{L}(V)$ is **invertible** if there exists another linear operator $S \in \mathcal{L}(V)$ such that ST and TS are identity maps (written as I). The linear operator S is called the **inverse** of T . The inverse is actually unique. Say S and S' are inverses of T . Then we have

$$S = SI = S(TS') = (ST)S' = IS' = S'. \quad (2.22)$$

Note that we required the inverse S to be an inverse acting from the left and acting from the right. This is useful for infinite dimensional vector spaces. For finite-dimensional vector spaces one suffices; one can then show that $ST = I$ if and only if $TS = I$.

It is useful to have a good characterization of invertible linear operators. For a finite-dimensional vector space V the following three statements are equivalent!

$$\text{Finite dimension: } \boxed{T \text{ is invertible}} \longleftrightarrow \boxed{T \text{ is injective}} \longleftrightarrow \boxed{T \text{ is surjective}} \quad (2.23)$$

For infinite dimensional vector spaces injectivity and surjectivity are not equivalent (each can fail independently). In that case invertibility is equivalent to injectivity plus surjectivity:

$$\text{Infinite dimension: } \boxed{T \text{ is invertible}} \longleftrightarrow \boxed{T \text{ is injective and surjective}} \quad (2.24)$$

The left shift operator L is not injective (maps $(x_1, 0, \dots)$ to zero) but it is surjective. The right shift operator is not surjective although it is injective.

Now we consider the **matrix associated to a linear operator T** that acts on a vector space V . This matrix will depend on the basis we choose for V . Let us declare that our basis is the list (v_1, v_2, \dots, v_n) . It is clear that the full knowledge of the action of T on V is encoded in the action of T on the basis vectors, that is on the values $(Tv_1, Tv_2, \dots, Tv_n)$. Since Tv_j is in V , it can be written as a linear combination of basis vectors. We then have

$$\boxed{Tv_j = T_{1j} v_1 + T_{2j} v_2 + \dots + T_{nj} v_n,} \quad (2.25)$$

where we introduced the constants T_{ij} that are known if the operator T is known. As we will see, these are the entries form the matrix representation of the operator T in the chosen basis. The above relation can be written more briefly as

$$\boxed{Tv_j = \sum_{i=1}^n T_{ij} v_i.} \quad (2.26)$$

When we deal with different bases it can be useful to use notation where we replace

$$T_{ij} \rightarrow T_{ij}(\{v\}), \quad (2.27)$$

so that it makes clear that T is being represented using the v basis (v_1, \dots, v_n) .

I want to make clear why (2.25) is reasonable before we show that it makes for a consistent association between operator multiplication and matrix multiplication. The left-hand side, where we have the action of the matrix for T on the j -th basis vector, can be viewed concretely as

$$Tv_j \longleftrightarrow \begin{pmatrix} T_{11} & \cdots & T_{1j} & \cdots & T_{1n} \\ T_{21} & \cdots & T_{2j} & \cdots & T_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{n1} & \cdots & T_{nj} & \cdots & T_{nn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad j\text{-th position} \quad (2.28)$$

where the column vector has zeroes everywhere except on the j -th entry. The product, by the usual rule of matrix multiplication is the column vector

$$\begin{pmatrix} T_{1j} \\ T_{2j} \\ \vdots \\ T_{nj} \end{pmatrix} = T_{1j} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + T_{2j} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \dots + T_{nj} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \longleftrightarrow T_{1j} v_1 + \dots + T_{nj} v_n. \quad (2.29)$$

which we identify with the right-hand side of (2.25). So (2.25) is reasonable.

Exercise. Verify that the matrix representation of the identity operator is a diagonal matrix with an entry of one at each element of the diagonal. This is true for any basis.

Let us now examine the product of two operators and their matrix representation. Consider the operator TS acting on v_j :

$$(TS)v_j = T(Sv_j) = T \sum_p S_{pj} v_p = \sum_p S_{pj} T v_p = \sum_p S_{pj} \sum_i T_{ip} v_i \quad (2.30)$$

so that changing the order of the sums we find

$$(TS)v_j = \sum_i \left(\sum_p T_{ip} S_{pj} \right) v_i. \quad (2.31)$$

Using the identification implicit in (2.26) we see that the object in parenthesis is the i, j matrix element of the matrix that represents TS . Therefore we found

$$(TS)_{ij} = \sum_p T_{ip} S_{pj}, \quad (2.32)$$

which is precisely the right formula for matrix multiplication. In other words, the matrix that represents TS is the product of the matrix that represents T with the matrix that represents S , in that order.

Changing basis

While matrix representations are very useful for concrete visualization, they are basis dependent. It is a good idea to try to figure out if there are quantities that can be calculated using a matrix representation that are, nevertheless, guaranteed to be basis independent. One such quantity is the **trace** of the matrix representation of a linear operator. The trace is the sum of the matrix elements in the diagonal. Remarkably, that sum is the same independent of the basis used. Consider a linear operator T in $\mathcal{L}(V)$ and two sets of basis vectors (v_1, \dots, v_n) and (u_1, \dots, u_n) for V . Using the explicit notation (2.27) for the matrix representation we state this property as

$$\text{tr } T(\{v\}) = \text{tr } T(\{u\}). \quad (2.33)$$

We will establish this result below. On the other hand, if this trace is actually basis independent, there should be a way to define the trace of the linear operator T *without* using its matrix representation. This is actually possible, as we will see. Another basis independent quantity is the determinant of the matrix representation of T .

Let us then consider the effect of a change of basis on the matrix representation of an operator. Consider a vector space V and a change of basis from (v_1, \dots, v_n) to (u_1, \dots, u_n) defined by the linear operator A as follows:

$$A : v_k \rightarrow u_k, \text{ for } k = 1, \dots, n. \quad (2.34)$$

This can also be written as

$$Av_k = u_k \quad (2.35)$$

Since we know how A acts on every element of the basis we know, by linearity how it acts on any vector. The operator A is clearly *invertible* because, letting $B : u_k \rightarrow v_k$ or

$$Bu_k = v_k, \quad (2.36)$$

we have

$$\begin{aligned} BA v_k &= B(Av_k) = Bu_k = v_k \\ AB u_k &= A(Bu_k) = Av_k = u_k, \end{aligned} \quad (2.37)$$

showing that $BA = I$ and $AB = I$. Thus B is the inverse of A . Using the definition of matrix representation, the right-hand sides of the relations $u_k = Av_k$ and $v_k = Bu_k$ can be written so that the equations take the form

$$u_k = A_{jk} v_j, \quad v_k = B_{jk} u_j, \quad (2.38)$$

where we used the convention that repeated indices are summed over. A_{ij} are the elements of the matrix representation of A in the v basis and B_{ij} are the elements of the matrix representation of B in the u basis. Replacing the second relation on the first, and then replacing the first on the second we get

$$\begin{aligned} u_k &= A_{jk} B_{ij} u_i = B_{ij} A_{jk} u_i \\ v_k &= B_{jk} A_{ij} v_i = A_{ij} B_{jk} v_i \end{aligned} \quad (2.39)$$

Since the u 's and v 's are basis vectors we must have

$$B_{ij} A_{jk} = \delta_{ik} \quad \text{and} \quad A_{ij} B_{jk} = \delta_{ik} \quad (2.40)$$

which means that the B matrix is the inverse of the A matrix. We have thus learned that

$$v_k = (A^{-1})_{jk} u_j. \quad (2.41)$$

We can now apply these preparatory results to the matrix representations of the operator T . We have, by definition,

$$Tv_k = T_{ik}(\{v\}) v_i. \quad (2.42)$$

We now want to calculate T on u_k so that we can read the formula for the matrix T on the u basis:

$$Tu_k = T_{ik}(\{u\}) u_i. \quad (2.43)$$

Computing the left-hand side, using the linearity of the operator T , we have

$$Tu_k = T(A_{jk} v_j) = A_{jk} T v_j = A_{jk} T_{pj}(\{v\}) v_p \quad (2.44)$$

and using (2.41) we get

$$Tu_k = A_{jk}T_{pj}(\{v\})(A^{-1})_{ip}u_i = \left((A^{-1})_{ip}T_{pj}(\{v\})A_{jk} \right)u_i = (A^{-1}T(\{v\})A)_{ik}u_i. \quad (2.45)$$

Comparing with (2.43) we get

$$T_{ij}(\{u\}) = (A^{-1}T(\{v\})A)_{ij} \rightarrow \boxed{T(\{u\}) = A^{-1}T(\{v\})A.} \quad (2.46)$$

This is the result we wanted to obtain.

The trace of a matrix T_{ij} is given by T_{ii} , where sum over i is understood. To show that the trace of T is basis independent we write

$$\begin{aligned} \text{tr}(T(\{u\})) &= T_{ii}(\{u\}) = (A^{-1})_{ij}T_{jk}(\{v\})A_{ki} \\ &= A_{ki}(A^{-1})_{ij}T_{jk}(\{v\}) \\ &= \delta_{kj}T_{jk}(\{v\}) = T_{jj}(\{v\}) = \text{tr}(T(\{v\})). \end{aligned} \quad (2.47)$$

For the determinant we recall that $\det(AB) = (\det A)(\det B)$. Therefore $\det(A)\det(A^{-1}) = 1$. From (2.46) we then get

$$\det T(\{u\}) = \det(A^{-1})\det T(\{v\})\det A = \det T(\{v\}). \quad (2.48)$$

Thus the determinant of the matrix that represents a linear operator is independent of the basis used.

3 Eigenvalues and eigenvectors

In quantum mechanics we need to consider eigenvalues and eigenstates of hermitian operators acting on complex vector spaces. These operators are called observables and their eigenvalues represent possible results of a measurement. In order to acquire a better perspective on these matters, we consider the eigenvalue/eigenvector problem in more generality.

One way to understand the action of an operator $T \in \mathcal{L}(V)$ on a vector space V is to understand how it acts on subspaces of V , as those are smaller than V and thus possibly simpler to deal with. Let U denote a subspace of V . In general, the action of T may take elements of U outside U . We have a noteworthy situation if T acting on any element of U gives an element of U . In this case U is said to be **invariant** under T , and T is then a well-defined linear operator on U . A very interesting situation arises if a suitable list of invariant subspaces give the space V as a direct sum.

Of all subspaces, one-dimensional ones are the simplest. Given some vector $u \in V$ one can consider the one-dimensional subspace U spanned by u :

$$U = \{cu : c \in \mathbb{F}\}. \quad (3.49)$$

We can ask if the one-dimensional subspace U is left invariant by the operator T . For this Tu must be equal to a number times u , as this guarantees that $Tu \in U$. Calling the number λ , we write

$$Tu = \lambda u. \quad (3.50)$$

This equation is so ubiquitous that names have been invented to label the objects involved. The number $\lambda \in \mathbb{F}$ is called an **eigenvalue** of the linear operator T if there is a **nonzero** vector $u \in V$ such that the equation above is satisfied. Suppose we find for some specific λ a nonzero vector u satisfying this equation. Then it follows that cu , for any $c \in \mathbb{F}$ also satisfies equation (3.50), so that the solution space of the equation includes the subspace U , which is now said to be an invariant subspace under T . It is convenient to call any vector that satisfies (3.50) for a given λ an **eigenvector** of T corresponding to λ . In doing so we are including the zero vector as a solution and thus as an eigenvector. It can often happen that for a given λ there are several linearly independent eigenvectors. In this case the invariant subspace associated with the eigenvalue λ is higher dimensional. The set of eigenvalues of T is called the **spectrum** of T .

Our equation above is equivalent to

$$(T - \lambda I) u = 0, \quad (3.51)$$

for some nonzero u . It is therefore the case that

λ is an eigenvalue	\longleftrightarrow	$(T - \lambda I)$ not injective.
----------------------------	-----------------------	----------------------------------

(3.52)

Using (2.23) we conclude that λ is an eigenvalue also means that $(T - \lambda I)$ is **not invertible**, and not surjective. We also note that

$$\text{Set of eigenvectors of } T \text{ corresponding to } \lambda = \text{null}(T - \lambda I). \quad (3.53)$$

It should be emphasized that the eigenvalues of T and the invariant subspaces (or eigenvectors associated with fixed eigenvalues) are basis independent objects. Nowhere in our discussion we had to invoke the use of a basis, nor we had to use any matrix representation. Below, we will discuss the familiar calculation of eigenvalues and eigenvectors using a matrix representation of the operator T in some particular basis.

Let us consider some examples. Take a real three-dimensional vector space V (our space to great accuracy!). Consider the rotation operator T that rotates all vectors by a fixed angle small about the z axis. To find eigenvalues and eigenvectors we just think of the invariant subspaces. We must ask which are the vectors for which this rotation doesn't change their direction and effectively just multiplies them by a number? Only the vectors along the z -direction do not change direction upon this rotation. So the vector space spanned by \mathbf{e}_z is the invariant subspace, or the space of eigenvectors. The eigenvectors are associated with the eigenvalue of one, as the vectors are not altered at all by the rotation.

Consider now the case where T is a rotation by ninety degrees on a two-dimensional *real* vector space V . Are there one-dimensional subspaces left invariant by T ? No, all vectors are rotated, none remains pointing in the same direction. Thus there are **no eigenvalues**, nor, of course, eigenvectors. If you tried calculating the eigenvalues by the usual recipe, you will find complex numbers. A complex eigenvalue is meaningless in a real vector space.

Although we will not prove the following result, it follows from the facts we have introduced and no extra machinery. It is of interest being completely general and valid for both real and complex vector spaces:

Theorem: Let $T \in \mathcal{L}(V)$ and assume $\lambda_1, \dots, \lambda_n$ are distinct eigenvalues of T and u_1, \dots, u_n are corresponding nonzero eigenvectors. Then (u_1, \dots, u_n) are linearly independent.

Note that we cannot ask if the eigenvectors are orthogonal to each other as we have not yet introduced an inner product on the vector space V . In this theorem there may be more than one linearly independent eigenvector associated with some eigenvalues. In that case any one eigenvector will do. Since an n -dimensional vector space V does not have more than n linearly independent vectors, no linear operator on V can have more than n distinct eigenvalues.

We saw that some linear operators in real vector spaces can fail to have eigenvalues. Complex vector spaces are nicer. In fact, *every linear operator on a finite-dimensional complex vector space has at least one eigenvalue*. This is a fundamental result. It can be proven without using determinants with an elegant argument, but the proof using determinants is quite short.

When λ is an eigenvalue, we have seen that $T - \lambda I$ is not an invertible operator. This also means that using any basis, the matrix representative of $T - \lambda I$ is non-invertible. The condition of non-invertibility of a matrix is identical to the condition that its determinant vanish:

$$\det(T - \lambda \mathbf{1}) = 0. \quad (3.54)$$

This condition, in an N -dimensional vector space looks like

$$\det \begin{pmatrix} T_{11} - \lambda & T_{12} & \dots & T_{1N} \\ T_{21} & T_{22} - \lambda & \dots & T_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ T_{N1} & T_{N2} & \dots & T_{NN} - \lambda \end{pmatrix} = 0. \quad (3.55)$$

The left-hand side is a polynomial $f(\lambda)$ in λ of degree N called the *characteristic polynomial*:

$$f(\lambda) = \det(T - \lambda \mathbf{1}) = (-\lambda)^N + b_{N-1}\lambda^{N-1} + \dots + b_1\lambda + b_0, \quad (3.56)$$

where the b_i are constants. We are interested in the equation $f(\lambda) = 0$, as this determines all possible eigenvalues. If we are working on real vector spaces, the constants b_i are real but there is no guarantee of real roots for $f(\lambda) = 0$. With complex vector spaces, the constants b_i will be complex, but a complex solution for $f(\lambda) = 0$ always exists. Indeed, over the complex numbers we can factor the polynomial $f(\lambda)$ as follows

$$f(\lambda) = (-1)^N (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_N), \quad (3.57)$$

where the notation does not preclude the possibility that some of the λ_i 's may be equal. The λ_i 's are the eigenvalues, since they lead to $f(\lambda) = 0$ for $\lambda = \lambda_i$. If all eigenvalues of T are different the spectrum of T is said to be *non-degenerate*. If an eigenvalue appears k times it is said to be a degenerate eigenvalue with of multiplicity k . Even in the most degenerate case we must have at least one eigenvalue. The eigenvectors exist because $(T - \lambda I)$ non-invertible means it is not injective, and therefore there are nonzero vectors that are mapped to zero by this operator.

4 Inner products

We have been able to go a long way without introducing extra structure on the vector spaces. We have considered linear operators, matrix representations, traces, invariant subspaces, eigenvalues and eigenvectors. It is now time to put some additional structure on the vector spaces. In this section we consider a function called an *inner product* that allows us to construct numbers from vectors. A vector space equipped with an inner product is called an inner-product space.

An **inner product** on a vector space V over \mathbb{F} is a machine that takes an *ordered* pair of elements of V , that is, a first vector and a second vector, and yields a number in \mathbb{F} . In order to motivate the definition of an inner product we first discuss the familiar way in which we associate a length to a vector.

The length of a vector, or **norm** of a vector is a real number that is positive or zero, if the vector is the zero vector. In \mathbb{R}^n a vector $a = (a_1, \dots, a_n)$ has norm $|a|$ defined by

$$|a| = \sqrt{a_1^2 + \dots + a_n^2} \quad (4.58)$$

Squaring this one may think of $|a|^2$ as the *dot product* of a with a :

$$|a|^2 = a \cdot a = a_1^2 + \dots + a_n^2 \quad (4.59)$$

Based on this the dot product of any two vectors a and b is defined by

$$a \cdot b = a_1 b_1 + \dots + a_n b_n. \quad (4.60)$$

If we try to generalize this dot product we may require as needed properties the following

1. $a \cdot a \geq 0$, for all vectors a .
2. $a \cdot a = 0$ if and only if $a = 0$.
3. $a \cdot (b_1 + b_2) = a \cdot b_1 + a \cdot b_2$. Additivity in the second entry.
4. $a \cdot (\alpha b) = \alpha a \cdot b$, with α a number.
5. $a \cdot b = b \cdot a$.

Along with these axioms, the length $|a|$ of a vector a is the positive or zero number defined by relation

$$|a|^2 = a \cdot a. \quad (4.61)$$

These axioms are satisfied by the definition (4.60) but do not require it. A new dot product defined by $a \cdot b = c_1 a_1 b_1 + \dots + c_n a_n b_n$, with c_1, \dots, c_n positive constants, would do equally well! So whatever can be proven with these axioms holds true not only for the conventional dot product.

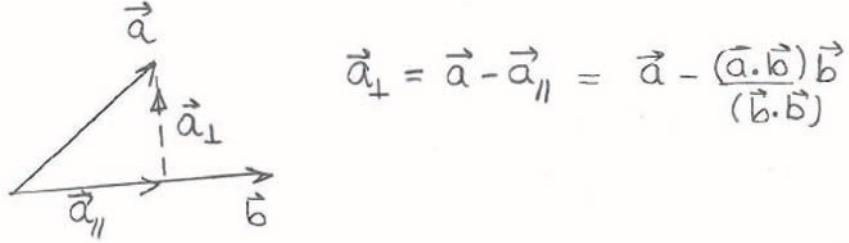
The above axioms guarantee that the Schwarz inequality holds:

$$|a \cdot b| \leq |a| |b|. \quad (4.62)$$

To prove this consider two (nonzero) vectors a and b and then consider the shortest vector joining the tip of a to the line defined by the direction of b (see the figure below). This is the vector a_{\perp} , given by

$$a_{\perp} \equiv a - \frac{a \cdot b}{b \cdot b} b. \quad (4.63)$$

The subscript \perp is there because the vector is perpendicular to b , namely $a_{\perp} \cdot b = 0$, as you can quickly see. To write the above vector we subtracted from a the component of a parallel to b . Note that the vector a_{\perp} is not changed as $b \rightarrow cb$; it does not depend on the overall length of b . Moreover, as it should, the vector a_{\perp} is zero if and only if the vectors a and b are parallel. All this is only motivation, we could have just said “consider the following vector a_{\perp} ”.



Given axiom (1) we have that $a_{\perp} \cdot a_{\perp} \geq 0$ and therefore using (4.63)

$$a_{\perp} \cdot a_{\perp} = a \cdot a - \frac{(a \cdot b)^2}{b \cdot b} \geq 0. \quad (4.64)$$

Since b is not the zero vector we then have

$$(a \cdot b)^2 \leq (a \cdot a)(b \cdot b). \quad (4.65)$$

Taking the square root of this relation we obtain the Schwarz inequality (4.62). The inequality becomes an equality only if $a_{\perp} = 0$ or, as discussed above, when $a = cb$ with c a real constant.

For complex vector spaces some modification is necessary. Recall that the length $|\gamma|$ of a complex number γ is given by $|\gamma| = \sqrt{\gamma^* \gamma}$, where the asterisk superscript denotes complex conjugation. It is

not hard to generalize this a bit. Let $z = (z_1, \dots, z_n)$ be a vector in \mathbb{C}^n . Then the length of the vector $|z|$ is a real number greater than zero given by

$$|z| = \sqrt{z_1^* z_1 + \dots + z_n^* z_n}. \quad (4.66)$$

We must use complex conjugates, denoted by the asterisk superscript, to produce a real number greater than or equal to zero. Squaring this we have

$$|z|^2 = z_1^* z_1 + \dots + z_n^* z_n. \quad (4.67)$$

This suggests that for vectors $z = (z_1, \dots, z_n)$ and $w = (w_1, \dots, w_n)$ an inner product could be given by

$$w_1^* z_1 + \dots + w_n^* z_n, \quad (4.68)$$

and we see that we are not treating the two vectors in an equivalent way. There is the first vector, in this case w whose components are conjugated and a second vector z whose components are not conjugated. If the order of vectors is reversed, we get for the inner product the complex conjugate of the original value. As it was mentioned at the beginning of the section, the inner product requires an ordered pair of vectors. It certainly does for complex vector spaces. Moreover, one can define an inner product in general in a way that applies both to complex and real vector spaces.

An **inner product** on a vector space V over \mathbb{F} is a map from an ordered pair (u, v) of vectors in V to a number $\langle u, v \rangle$ in \mathbb{F} . The axioms for $\langle u, v \rangle$ are inspired by the axioms we listed for the dot product.

1. $\langle v, v \rangle \geq 0$, for all vectors $v \in V$.
2. $\langle v, v \rangle = 0$ if and only if $v = 0$.
3. $\langle u, v_1 + v_2 \rangle = \langle u, v_1 \rangle + \langle u, v_2 \rangle$. Additivity in the second entry.
4. $\langle u, \alpha v \rangle = \alpha \langle u, v \rangle$, with $\alpha \in \mathbb{F}$. Homogeneity in the second entry.
5. $\langle u, v \rangle = \langle v, u \rangle^*$. Conjugate exchange symmetry.

This time the **norm** $|v|$ of a vector $v \in V$ is the positive or zero number defined by relation

$$|v|^2 = \langle v, v \rangle. \quad (4.69)$$

From the axioms above, the only major difference is in number five, where we find that the inner product is not symmetric. We know what complex conjugation is in \mathbb{C} . For the above axioms to apply to vector spaces over \mathbb{R} we just define the obvious: complex conjugation of a real number is a real number. In a real vector space the * conjugation does nothing and the inner product is strictly symmetric in its inputs.

A few comments. One can use (3) with $v_2 = 0$ to show that $\langle u, 0 \rangle = 0$ for all $u \in V$, and thus, by (5) also $\langle 0, u \rangle = 0$. Properties (3) and (4) amount to full linearity in the second entry. It is important to note that additivity holds for the first entry as well:

$$\begin{aligned}\langle u_1 + u_2, v \rangle &= \langle v, u_1 + u_2 \rangle^* \\ &= (\langle v, u_1 \rangle + \langle v, u_2 \rangle)^* \\ &= \langle v, u_1 \rangle^* + \langle v, u_2 \rangle^* \\ &= \langle u_1, v \rangle + \langle u_2, v \rangle.\end{aligned}\tag{4.70}$$

Homogeneity works differently on the first entry, however,

$$\begin{aligned}\langle \alpha u, v \rangle &= \langle v, \alpha u \rangle^* \\ &= (\alpha \langle v, u \rangle)^* \\ &= \alpha^* \langle u, v \rangle.\end{aligned}\tag{4.71}$$

Thus we get **conjugate homogeneity** on the first entry. This is a very important fact. Of course, for a real vector space conjugate homogeneity is the same as just plain homogeneity.

Two vectors $u, v \in V$ are said to be **orthogonal** if $\langle u, v \rangle = 0$. This, of course, means that $\langle v, u \rangle = 0$ as well. The zero vector is orthogonal to all vectors (including itself). Any vector orthogonal to all vectors in the vector space must be equal to zero. Indeed, if $x \in V$ is such that $\langle x, v \rangle = 0$ for all v , pick $v = x$, so that $\langle x, x \rangle = 0$ implies $x = 0$ by axiom 2. This property is sometimes stated as the **non-degeneracy** of the inner product. The “Pythagorean” identity holds for the norm-squared of orthogonal vectors in an inner-product vector space. As you can quickly verify,

$$|u + v|^2 = |u|^2 + |v|^2, \quad \text{for } u, v \in V, \text{ orthogonal vectors.}\tag{4.72}$$

The Schwarz inequality can be proven by an argument fairly analogous to the one we gave above for dot products. The result now reads

Schwarz Inequality: $|\langle u, v \rangle| \leq |u| |v|.$

(4.73)

The inequality is saturated if and only if one vector is a multiple of the other. Note that in the left-hand side $|...|$ denotes the norm of a complex number and on the right-hand side each $|...|$ denotes the norm of a vector. You will prove this identity in a slightly different way in the homework. You will also consider there the *triangle inequality*

$$|u + v| \leq |u| + |v|,\tag{4.74}$$

which is saturated when $u = cv$ for c a real, positive constant. Our definition (4.69) of norm on a vector space V is mathematically sound: a norm is required to satisfy the triangle inequality. Other properties are required: (i) $|v| \geq 0$ for all v , (ii) $|v| = 0$ if and only if $v = 0$, and (iii) $|cv| = |c||v|$ for c some constant. Our norm satisfies all of them.

A complex vector space with an inner product as we have defined is a *Hilbert space* if it is finite dimensional. If the vector space is infinite dimensional, an extra *completeness* requirement must be satisfied for the space to be a Hilbert space: all Cauchy sequences of vectors must converge to vectors in the space. An infinite sequence of vectors v_i , with $i = 1, 2, \dots, \infty$ is a Cauchy sequence if for any $\epsilon > 0$ there is an N such that $|v_n - v_m| < \epsilon$ whenever $n, m > N$.

5 Orthonormal basis and orthogonal projectors

In an inner-product space we can demand that basis vectors have special properties. A list of vectors is said to be **orthonormal** if all vectors have norm one and are pairwise orthogonal. Consider a list (e_1, \dots, e_n) of orthonormal vectors in V . Orthonormality means that

$$\langle e_i, e_j \rangle = \delta_{ij}. \quad (5.75)$$

We also have a simple expression for the norm of $a_1e_1 + \dots + a_ne_n$, with $a_i \in \mathbb{F}$:

$$\begin{aligned} |a_1e_1 + \dots + a_ne_n|^2 &= \langle a_1e_1 + \dots + a_ne_n, a_1e_1 + \dots + a_ne_n \rangle \\ &= \langle a_1e_1, a_1e_1 \rangle + \dots + \langle a_ne_n, a_ne_n \rangle \\ &= |a_1|^2 + \dots + |a_n|^2. \end{aligned} \quad (5.76)$$

This result implies the somewhat nontrivial fact that *the vectors in any orthonormal list are linearly independent*. Indeed if $a_1e_1 + \dots + a_ne_n = 0$ then its norm is zero and so is $|a_1|^2 + \dots + |a_n|^2$. This implies all $a_i = 0$, thus proving the claim.

An **orthonormal basis** of V is a list of orthonormal vectors that is also a basis for V . Let (e_1, \dots, e_n) denote an orthonormal basis. Then any vector v can be written as

$$v = a_1e_1 + \dots + a_ne_n, \quad (5.77)$$

for some constants a_i that can be calculated as follows

$$\langle e_i, v \rangle = \langle e_i, a_1e_1 + \dots + a_ne_n \rangle = a_i, \quad (i \text{ not summed}). \quad (5.78)$$

Therefore any vector v can be written as

$$v = \langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n = \langle e_i, v \rangle e_i. \quad (5.79)$$

To find an orthonormal basis on an inner product space V we just need to start with a basis and then use an algorithm to turn it into an orthogonal basis. In fact, a little more generally:

Gram-Schmidt: Given a list (v_1, \dots, v_n) of linearly independent vectors in V one can construct a list (e_1, \dots, e_n) of orthonormal vectors such that both lists span the same subspace of V .

The Gram-Schmidt algorithm goes as follows. You take e_1 to be v_1 , normalized to have unit norm: $e_1 = v_1/|v_1|$. Then take $v_2 + \alpha e_1$ and fix the constant α so that this vector is orthogonal to e_1 . The

answer is clearly $v_2 - \langle e_1, v_2 \rangle e_1$. This vector, normalized by dividing it by its norm, is set equal to e_2 . In fact we can write the general vector in a recursive fashion. If we know e_1, e_2, \dots, e_{j-1} , we can write e_j as follows:

$$e_j = \frac{v_j - \langle e_1, v_j \rangle e_1 - \dots - \langle e_{j-1}, v_j \rangle e_{j-1}}{|v_j - \langle e_1, v_j \rangle e_1 - \dots - \langle e_{j-1}, v_j \rangle e_{j-1}|} \quad (5.80)$$

It should be clear to you by inspection that this vector is orthogonal to the vectors e_i with $i < j$ and has unit norm. The Gram-Schmidt procedure is quite practical.

With an inner product we can construct interesting subspaces of a vector space V . Consider a subset U of vectors in V (not necessarily a subspace). Then we can define a subspace U^\perp , called the **orthogonal complement** of U as the set of all vectors orthogonal to the vectors in U :

$$U^\perp = \{v \in V \mid \langle v, u \rangle = 0, \text{ for all } u \in U\}. \quad (5.81)$$

This is clearly a subspace of V . When U is a subspace, then U and U^\perp actually give a direct sum decomposition of the full space:

Theorem: If U is a subspace of V , then $V = U \oplus U^\perp$.

Proof: This is a fundamental result and is not hard to prove. Let (e_1, \dots, e_n) be an orthonormal basis for U . We can clearly write any vector v in V as

$$v = (\langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n) + (v - \langle e_1, v \rangle e_1 - \dots - \langle e_n, v \rangle e_n). \quad (5.82)$$

On the right-hand side the first vector in parenthesis is clearly in U as it is written as a linear combination of U basis vectors. The second vector is clearly in U^\perp as one can see that it is orthogonal to any vector in U . To complete the proof one must show that there is no vector except the zero vector in the intersection $U \cap U^\perp$ (recall the comments below (1.5)). Let $v \in U \cap U^\perp$. Then v is in U and in U^\perp so it should satisfy $\langle v, v \rangle = 0$. But then $v = 0$, completing the proof.

Given this decomposition any vector $v \in V$ can be written uniquely as $v = u + w$ where $u \in U$ and $w \in U^\perp$. One can define a linear operator P_U , called the **orthogonal projection** of V onto U , that acting on v above gives the vector u . It is clear from this definition that: (i) the range of P_U is U , (ii) the null space of P_U is U^\perp , (iii) that P_U is not invertible and, (iv) acting on U , the operator P_U is the identity operator. The formula for the vector u can be read from (5.82)

$$P_U v = \langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n. \quad (5.83)$$

It is a straightforward but a good exercise to verify that this formula is consistent with the fact that acting on U , the operator P_U is the identity operator. Thus if we act twice in succession with P_U on a vector, the second action has no effect as it is already acting on a vector in U . It follows from this that

$$P_U P_U = I P_U = P_U \rightarrow \boxed{P_U^2 = P_U}. \quad (5.84)$$

The eigenvalues and eigenvectors of P_U are easy to describe. Since all vectors in U are left invariant by the action of P_U , an orthonormal basis of U provides a set of orthonormal eigenvectors of P all with

eigenvalue one. If we choose on U^\perp an orthonormal basis, that basis provides orthonormal eigenvectors of P all with eigenvalue zero.

In fact equation (5.84) implies that the eigenvalues of P_U can only be one or zero. The eigenvalues of an operator satisfy whatever equation the operator satisfies (as shown by letting the equation act on a presumed eigenvector) thus $\lambda^2 = \lambda$ is needed, and this gives $\lambda(\lambda - 1) = 0$, and $\lambda = 0, 1$, as the only possibilities.

Consider a vector space $V = U \oplus U^\perp$ that is $(n + k)$ -dimensional, where U is n -dimensional and U^\perp is k -dimensional. Let (e_1, \dots, e_n) be an orthonormal basis for U and (f_1, \dots, f_k) an orthonormal basis for U^\perp . We then see that the list of vectors (g_1, \dots, g_{n+k}) defined by

$$(g_1, \dots, g_{n+k}) = (e_1, \dots, e_n, f_1, \dots, f_k) \text{ is an orthonormal basis for } V. \quad (5.85)$$

Exercise: Use $P_U e_i = e_i$, for $i = 1, \dots, n$ and $P_U f_i = 0$, for $i = 1, \dots, k$, to show that in the above basis the projector operator is represented by the diagonal matrix:

$$P_U = \text{diag}\left(\underbrace{1, \dots, 1}_{n \text{ entries}}, \underbrace{0, \dots, 0}_{k \text{ entries}}\right). \quad (5.86)$$

We see that, as expected from its non-invertibility, $\det(P_U) = 0$. But more interestingly we see that the trace of the matrix P_U is n . Therefore

$$\text{tr } P_U = \dim U. \quad (5.87)$$

The dimension of U is the **rank** of the projector P_U . Rank one projectors are the most common projectors. They project to one-dimensional subspaces of the vector space.

Projection operators are useful in quantum mechanics, where observables are described by operators. The effect of measuring an observable on a physical state vector is to turn this original vector instantaneously into another vector. This resulting vector is the orthogonal projection of the original vector down to some eigenspace of the operator associated with the observable.

6 Linear functionals and adjoint operators

When we consider a linear operator T on a vector space V that has an inner product, we can construct a related linear operator T^\dagger on V called the **adjoint** of T . This is a very useful operator and is typically different from T . When the adjoint T^\dagger happens to be equal to T , the operator is said to be *Hermitian*. To understand adjoints, we first need to develop the concept of a linear functional.

A **linear functional** ϕ on the vector space V is a linear map from V to the numbers \mathbb{F} : for $v \in V$, $\phi(v) \in \mathbb{F}$. A linear functional has the following two properties:

1. $\phi(v_1 + v_2) = \phi(v_1) + \phi(v_2)$, with $v_1, v_2 \in V$.
2. $\phi(av) = a\phi(v)$ for $v \in V$ and $a \in \mathbb{F}$.

As an example, consider the three-dimensional real vector space \mathbb{R}^3 with inner product equal to the familiar dot product. Writing a vector v as the triplet $v = (v_1, v_2, v_3)$, we take

$$\phi(v) = 3v_1 + 2v_2 - 4v_3. \quad (6.1)$$

Linearity is clear as the right-hand side features the components v_1, v_2, v_3 appearing linearly. We can use a vector $u = (3, 2, -4)$ to write the linear functional as an inner product. Indeed, one can readily see that

$$\phi(v) = \langle u, v \rangle. \quad (6.2)$$

This is no accident, in fact. We can prove that any linear functional $\phi(v)$ admits such representation with some suitable choice of vector u .

Theorem: Let ϕ be a linear functional on V . There is a unique vector $u \in V$ such that $\phi(v) = \langle u, v \rangle$ for all $v \in V$.

Proof: Consider an orthonormal basis, (e_1, \dots, e_n) and write the vector v as

$$v = \langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n. \quad (6.3)$$

When ϕ acts on v we find, first by linearity and then by conjugate homogeneity

$$\begin{aligned} \phi(v) &= \phi(\langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n) \\ &= \langle e_1, v \rangle \phi(e_1) + \dots + \langle e_n, v \rangle \phi(e_n) \\ &= \langle \phi(e_1)^* e_1, v \rangle + \dots + \langle \phi(e_n)^* e_n, v \rangle \\ &= \langle \phi(e_1)^* e_1 + \dots + \phi(e_n)^* e_n, v \rangle. \end{aligned} \quad (6.4)$$

We have thus shown that, as claimed

$$\phi(v) = \langle u, v \rangle \quad \text{with} \quad u = \phi(e_1)^* e_1 + \dots + \phi(e_n)^* e_n. \quad (6.5)$$

Next, we prove that this u is unique. If there exists another vector, u' , that also gives the correct result for all v , then $\langle u', v \rangle = \langle u, v \rangle$, which implies $\langle u - u', v \rangle = 0$ for all v . Taking $v = u' - u$, we see that this shows $u' - u = 0$ or $u' = u$, proving uniqueness.¹

We can modify a bit the notation when needed, to write

$$\phi_u(v) \equiv \langle u, v \rangle, \quad (6.6)$$

where the left-hand side makes it clear that this is a functional acting on v that depends on u .

We can now address the construction of the adjoint. Consider: $\phi(v) = \langle u, T v \rangle$, which is clearly a linear functional, whatever the operator T is. Since any linear functional can be written as $\langle w, v \rangle$, with some suitable vector w , we write

$$\langle u, T v \rangle = \langle w, v \rangle, \quad (6.7)$$

¹This theorem holds for infinite dimensional Hilbert spaces, for *continuous* linear functionals.

Of course, the vector w must depend on the vector u that appears on the left-hand side. Moreover, it must have something to do with the operator T , who does not appear anymore on the right-hand side. So we must look for some good notation here. We can think of w as a function of the vector u and thus write $w = T^\dagger u$ where T^\dagger denotes a map (not obviously linear) from V to V . So, we think of $T^\dagger u$ as the vector obtained by acting with some function T^\dagger on u . The above equation is written as

$$\langle u, Tv \rangle = \langle T^\dagger u, v \rangle, \quad (6.8)$$

Our next step is to show that, in fact, T^\dagger is a linear operator on V . The operator T^\dagger is called the **adjoint** of T . Consider

$$\langle u_1 + u_2, Tv \rangle = \langle T^\dagger(u_1 + u_2), v \rangle, \quad (6.9)$$

and work on the left-hand side to get

$$\begin{aligned} \langle u_1 + u_2, Tv \rangle &= \langle u_1, Tv \rangle + \langle u_2, Tv \rangle \\ &= \langle T^\dagger u_1, v \rangle + \langle T^\dagger u_2, v \rangle \\ &= \langle T^\dagger u_1 + T^\dagger u_2, v \rangle. \end{aligned} \quad (6.10)$$

Comparing the right-hand sides of the last two equations we get the desired:

$$T^\dagger(u_1 + u_2) = T^\dagger u_1 + T^\dagger u_2. \quad (6.11)$$

Having established linearity now we establish homogeneity. Consider

$$\langle au, Tv \rangle = \langle T^\dagger(au), v \rangle. \quad (6.12)$$

The left hand side is

$$\langle au, Tv \rangle = a^* \langle u, Tv \rangle = a^* \langle T^\dagger u, v \rangle = \langle aT^\dagger u, v \rangle. \quad (6.13)$$

This time we conclude that

$$T^\dagger(au) = aT^\dagger u. \quad (6.14)$$

This concludes the proof that T^\dagger , so defined is a linear operator on V .

A couple of important properties are readily proven:

Claim: $(ST)^\dagger = T^\dagger S^\dagger$. We can show this as follows: $\langle u, STv \rangle = \langle S^\dagger u, Tv \rangle = \langle T^\dagger S^\dagger u, v \rangle$.

Claim: The adjoint of the adjoint is the original operator: $(S^\dagger)^\dagger = S$. We can show this as follows: $\langle u, S^\dagger v \rangle = \langle (S^\dagger)^\dagger u, v \rangle$. Now, additionally $\langle u, S^\dagger v \rangle = \langle S^\dagger v, u \rangle^* = \langle v, Su \rangle^* = \langle Su, v \rangle$. Comparing with the first result, we have shown that $(S^\dagger)^\dagger u = Su$, for any u , which proves the claim

Example: Let $v = (v_1, v_2, v_3)$, with $v_i \in \mathbb{C}$ denote a vector in the three-dimensional complex vector space, \mathbb{C}^3 . Define a linear operator T that acts on v as follows:

$$T(v_1, v_2, v_3) = (0v_1 + 2v_2 + iv_3, v_1 - iv_2 + 0v_3, 3iv_1 + v_2 + 7v_3). \quad (6.15)$$

Calculate the action of T^\dagger on a vector. Give the matrix representations of T and T^\dagger using the orthonormal basis $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$. Assume the inner product is the standard one on \mathbb{C}^3 .

Solution: We introduce a vector $u = (u_1, u_2, u_3)$ and will use the basic identity $\langle u, Tv \rangle = \langle T^\dagger u, v \rangle$. The left-hand side of the identity gives:

$$\langle u, Tv \rangle = u_1^*(2v_2 + iv_3) + u_2^*(v_1 - iv_2) + u_3^*(3iv_1 + v_2 + 7v_3). \quad (6.16)$$

This is now rewritten by factoring the various v_i 's

$$\langle u, Tv \rangle = (u_2^* + 3iu_3)v_1 + (2u_1^* - iu_2^* + u_3^*)v_2 + (iu_1^* + 7u_3^*)v_3. \quad (6.17)$$

Identifying the right-hand side with $\langle T^\dagger u, v \rangle$ we now deduce that

$$T^\dagger(u_1, u_2, u_3) = (u_2 - 3iu_3, 2u_1 + iu_2 + u_3, -iu_1 + 7u_3). \quad (6.18)$$

This gives the action of T^\dagger . To find the matrix representation we begin with T . Using basis vectors, we have from (6.15)

$$Te_1 = T(1, 0, 0) = (0, 1, 3i) = e_2 + 3ie_3 = T_{11}e_1 + T_{21}e_2 + T_{31}e_3, \quad (6.19)$$

and deduce that $T_{11} = 0$, $T_{21} = 1$, $T_{31} = 3i$. This can be repeated, and the rule becomes clear quickly: the coefficients of v_i read left to right fit into the i -th column of the matrix. Thus, we have

$$T = \begin{pmatrix} 0 & 2 & i \\ 1 & -i & 0 \\ 3i & 1 & 7 \end{pmatrix} \quad \text{and} \quad T^\dagger = \begin{pmatrix} 0 & 1 & -3i \\ 2 & i & 1 \\ -i & 0 & 7 \end{pmatrix}. \quad (6.20)$$

These matrices are related: one is the transpose and complex conjugate of the other! This is not an accident.

Let us reframe this using matrix notation. Let $u = e_i$ and $v = e_j$ where e_i and e_j are orthonormal basis vectors. Then the definition $\langle u, Tv \rangle = \langle T^\dagger u, v \rangle$ can be written as

$$\begin{aligned} \langle T^\dagger e_i, e_j \rangle &= \langle e_i, Te_j \rangle \\ \langle T_{ki}^\dagger e_k, e_j \rangle &= \langle e_i, T_{kj} e_k \rangle \\ (T_{ki}^\dagger)^* \delta_{kj} &= T_{jk} \delta_{ik} \\ (T^\dagger)_{ji}^* &= T_{ij} \end{aligned} \quad (6.21)$$

Relabeling i and j and taking the complex conjugate we find the familiar relation between a matrix and its adjoint:

$$(T^\dagger)_{ij} = (T_{ji})^*. \quad (6.22)$$

The adjoint matrix is the transpose and complex conjugate matrix only if we use an orthonormal basis. If we did not, in the equation above the use of $\langle e_i, e_j \rangle = \delta_{ij}$ would be replaced by $\langle e_i, e_j \rangle = g_{ij}$, where g_{ij} is some constant matrix that would appear in the rule for the construction of the adjoint matrix.

7 Hermitian and Unitary operators

Before we begin looking at special kinds of operators let us consider a very surprising fact about operators on complex vector spaces, as opposed to operators on real vector spaces.

Suppose we have an operator T that is such that for any vector $v \in V$ the following inner product vanishes

$$\langle v, T v \rangle = 0 \quad \text{for all } v \in V. \quad (7.23)$$

What can we say about the operator T ? The condition states that T is an operator that starting from a vector gives a vector orthogonal to the original one. In a two-dimensional real vector space, this is simply the operator that rotates any vector by ninety degrees! It is quite surprising and important that for *complex* vector spaces the result is very strong: any such operator T necessarily vanishes. This is a theorem:

Theorem: Let T be a linear operator in a **complex vector space** V :

$\text{If } \langle v, T v \rangle = 0 \text{ for all } v \in V, \text{ then } T = 0.$

(7.24)

Proof: Any proof must be such that it fails to work for real vector space. Note that the result follows if we could prove that $\langle u, T v \rangle = 0$, for all $u, v \in V$. Indeed, if this holds, then take $u = T v$, then $\langle T v, T v \rangle = 0$ for all v implies that $T v = 0$ for all v and therefore $T = 0$.

We will thus try to show that $\langle u, T v \rangle = 0$ for all $u, v \in V$. All we know is that objects of the form $\langle \#, T \# \rangle$ vanish, whatever $\#$ is. So we must aim to form linear combinations of such terms in order to reproduce $\langle u, T v \rangle$. We begin by trying the following

$$\langle u + v, T(u + v) \rangle - \langle u - v, T(u - v) \rangle = 2\langle u, T v \rangle + 2\langle v, T u \rangle. \quad (7.25)$$

We see that the “diagonal” term vanished, but instead of getting just $\langle u, T v \rangle$ we also got $\langle v, T u \rangle$. Here is where complex numbers help, we can get the same two terms but with opposite signs by trying,

$$\langle u + iv, T(u + iv) \rangle - \langle u - iv, T(u - iv) \rangle = 2i\langle u, T v \rangle - 2i\langle v, T u \rangle. \quad (7.26)$$

It follows from the last two relations that

$$\langle u, T v \rangle = \frac{1}{4} \left(\langle u + v, T(u + v) \rangle - \langle u - v, T(u - v) \rangle + \frac{1}{i} \langle u + iv, T(u + iv) \rangle - \frac{1}{i} \langle u - iv, T(u - iv) \rangle \right). \quad (7.27)$$

The condition $\langle v, T v \rangle = 0$ for all v , implies that each term of the above right-hand side vanishes, thus showing that $\langle u, T v \rangle = 0$ for all $u, v \in V$. As explained above this proves the result.

An operator T is said to be **Hermitian** if $T^\dagger = T$. Hermitian operators are pervasive in quantum mechanics. The above theorem in fact helps us discover Hermitian operators. It is familiar that the expectation value of a Hermitian operator, on any state, is real. It is also true, however, that any operator whose expectation value is real for all states must be Hermitian:

$$T = T^\dagger \text{ if and only if } \langle v, Tv \rangle \in \mathbb{R} \text{ for all } v. \quad (7.28)$$

To prove this first go from left to right. If $T = T^\dagger$

$$\langle v, Tv \rangle = \langle T^\dagger v, v \rangle = \langle Tv, v \rangle = \langle v, Tv \rangle^*, \quad (7.29)$$

showing that $\langle v, Tv \rangle$ is real. To go from right to left first note that the reality condition means that

$$\langle v, Tv \rangle = \langle Tv, v \rangle = \langle v, T^\dagger v \rangle, \quad (7.30)$$

where the last equality follows because $(T^\dagger)^\dagger = T$. Now the leftmost and rightmost terms can be combined to give $\langle v, (T - T^\dagger)v \rangle = 0$, which holding for all v implies, by the theorem, that $T = T^\dagger$.

We can prove two additional results of Hermitian operators rather easily. We have discussed earlier the fact that on a complex vector space any linear operator has at least one eigenvalue. Here we learn that the eigenvalues of a hermitian operator are real numbers. Moreover, while we have noted that eigenvectors corresponding to different eigenvalues are linearly independent, for Hermitian operators they are guaranteed to be orthogonal. Thus we have the following theorems

Theorem 1: The eigenvalues of Hermitian operators are real.

Theorem 2: Different eigenvalues of a Hermitian operator correspond to orthogonal eigenfunctions.

Proof 1: Let v be a nonzero eigenvector of the Hermitian operator T with eigenvalue λ : $Tv = \lambda v$. Taking the inner product with v we have that

$$\langle v, Tv \rangle = \langle v, \lambda v \rangle = \lambda \langle v, v \rangle. \quad (7.31)$$

Since T is hermitian, we can also evaluate $\langle v, Tv \rangle$ as follows

$$\langle v, Tv \rangle = \langle Tv, v \rangle = \langle \lambda v, v \rangle = \lambda^* \langle v, v \rangle. \quad (7.32)$$

The above equations give $(\lambda - \lambda^*) \langle v, v \rangle = 0$ and since v is not the zero vector, we conclude that $\lambda^* = \lambda$, showing the reality of λ .

Proof 2: Let v_1 and v_2 be eigenvectors of the operator T :

$$Tv_1 = \lambda_1 v_1, \quad Tv_2 = \lambda_2 v_2, \quad (7.33)$$

with λ_1 and λ_2 real (previous theorem) and different from each other. Consider the inner product $\langle v_2, Tv_1 \rangle$ and evaluate it in two different ways. First

$$\langle v_2, Tv_1 \rangle = \langle v_2, \lambda_1 v_1 \rangle = \lambda_1 \langle v_2, v_1 \rangle, \quad (7.34)$$

and second, using hermiticity of T ,

$$\langle v_2, T v_1 \rangle = \langle T v_2, v_1 \rangle = \langle \lambda_2 v_2, v_1 \rangle = \lambda_2 \langle v_2, v_1 \rangle. \quad (7.35)$$

From these two evaluations we conclude that

$$(\lambda_1 - \lambda_2) \langle v_1, v_2 \rangle = 0 \quad (7.36)$$

and the assumption $\lambda_1 \neq \lambda_2$, leads to $\langle v_1, v_2 \rangle = 0$, showing the orthogonality of the eigenvectors.

Let us now consider another important class of linear operators on a complex vector space, the so-called unitary operators. An operator $U \in \mathcal{L}(V)$ in a complex vector space V is said to be a **unitary operator** if it is surjective and does not change the magnitude of the vector it acts upon:

$$|Uu| = |u|, \text{ for all } u \in V. \quad (7.37)$$

We tailored the definition to be useful even for infinite dimensional spaces. Note that U can only kill vectors of zero length, and since the only such vector is the zero vector, $\text{null } U = 0$, and U is injective. Since U is also assumed to be surjective, a **unitary operator U is always invertible**.

A simple example of a unitary operator is the operator λI with λ a complex number of unit-norm: $|\lambda| = 1$. Indeed $|\lambda Iu| = |\lambda u| = |\lambda||u| = |u|$ for all u . Moreover, the operator is clearly surjective.

For another useful characterization of unitary operators we begin by squaring (7.37)

$$\langle Uu, Uu \rangle = \langle u, u \rangle \quad (7.38)$$

By the definition of adjoint

$$\langle u, U^\dagger U u \rangle = \langle u, u \rangle \rightarrow \langle u, (U^\dagger U - I)u \rangle = 0 \text{ for all } u. \quad (7.39)$$

So by our theorem $U^\dagger U = I$, and since U is invertible this means U^\dagger is the inverse of U and we also have $UU^\dagger = I$:

$$U^\dagger U = UU^\dagger = I.$$

(7.40)

Unitary operators *preserve inner products* in the following sense

$$\langle Uu, Uv \rangle = \langle u, v \rangle. \quad (7.41)$$

This follows immediately by moving the second U to act on the first input and using $U^\dagger U = I$.

Assume the vector space V is finite dimensional and has an orthonormal basis (e_1, \dots, e_n) . Consider the new set of vectors (f_1, \dots, f_n) where the f 's are obtained from the e 's by the action of a unitary operator U :

$$f_i = U e_i. \quad (7.42)$$

This also means that $e_i = U^\dagger f_i$. We readily see that the f 's are also a basis, because they are linearly independent: Acting on $a_1 f_1 + \dots + a_n f_n = 0$ with U^\dagger we find $a_1 e_1 + \dots + a_n e_n = 0$, and thus $a_i = 0$. We now see that the new basis is also orthonormal:

$$\langle f_i, f_j \rangle = \langle U e_i, U e_j \rangle = \langle e_i, e_j \rangle = \delta_{ij}. \quad (7.43)$$

The matrix elements of U in the e -basis are

$$U_{ki} = \langle e_k, U e_i \rangle. \quad (7.44)$$

Let us compute the matrix elements U'_{ki} of U in the f -basis

$$U'_{ki} = \langle f_k, U f_i \rangle = \langle U e_k, U f_i \rangle = \langle e_k, f_i \rangle = \langle e_k, U e_i \rangle = U_{ki} \quad (7.45)$$

The matrix elements are the same! Can you find an explanation for this result?

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 4

B. Zwiebach
February 18, 2016

Contents

1 de Broglie wavelength and Galilean transformations	1
2 Phase and Group Velocities	4
3 Choosing the wavefunction for a free particle	6

1 de Broglie wavelength and Galilean transformations

We have seen that to any free particle with momentum \mathbf{p} , we can associate a plane wave, or a “matter wave”, with de Broglie wavelength $\lambda = h/p$, with $p = |\mathbf{p}|$. The question is, *waves of what?* Well, this wave is eventually recognized as an example of what one calls the *wavefunction*. The wavefunction, as we will see is governed by the Schrödinger equation. As we have hinted, the wavefunction gives us information about probabilities, and we will develop this idea in detail.

Does the wave have directional or polarization properties like electric and magnetic fields in an electromagnetic wave? Yes, there is an analog of this, although we will not delve into it now. The analog of polarization corresponds to spin! The effects of spin are negligible in many cases (small velocities, no magnetic fields, for example) and for this reason, we just use a scalar wave, a complex number

$$\Psi(\mathbf{x}, t) \in \mathbb{C} \quad (1.1)$$

that depends on space and time. A couple of obvious questions come to mind. Is the wavefunction measurable? What kind of object is it? What does it describe? In trying to get intuition about this, let's consider how different observers perceive the de Broglie wavelength of a particle, which should help us understand what kind of waves we are talking about. Recall that

$$p = \frac{h}{\lambda} = \frac{h}{2\pi} \frac{2\pi}{\lambda} = \hbar k, \quad (1.2)$$

where k is the wavenumber. How would this wave behave under a change of frame?

We therefore consider two frames S and S' with the x and x' axes aligned and with S' moving to the right along the $+x$ direction of S with constant velocity v . At time equal zero, the origins of the two reference frames coincide.

The time and spatial coordinates of the two frames are related by a *Galilean transformation*, which states that

$$x' = x - vt, \quad t' = t. \quad (1.3)$$

Indeed time runs at the same speed in all Galilean frames and the relation between x and x' is manifest from the arrangement shown in Fig. 1.

Now assume both observers focus on a particle of mass m moving with nonrelativistic speed. Call the speed and momentum in the S frame \tilde{v} and $p = m\tilde{v}$, respectively. It follows by differentiation with

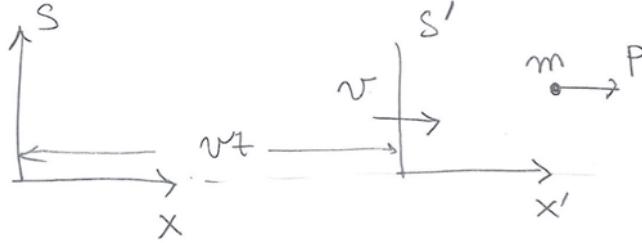


Figure 1: The S' frame moves at speed v along the x -direction of the S frame. A particle of mass m moves with speed \tilde{v} , and thus momentum $p = m\tilde{v}$, in the S frame.

respect to $t = t'$ of the first equation in (1.3) that

$$\frac{dx'}{dt'} = \frac{dx}{dt} - v, \quad (1.4)$$

which means that the particle velocity \tilde{v}' in the S' frame is given by

$$\tilde{v}' = \tilde{v} - v. \quad (1.5)$$

Multiplying by the mass m we find the relation between the momenta in the two frames

$$p' = p - mv. \quad (1.6)$$

The momentum p' in the S' frame can be appreciably different from the momentum p in the S frame. Thus the observers in S' and in S will obtain rather different de Broglie wavelengths λ' and λ ! Indeed,

$$\lambda' = \frac{h}{p'} = \frac{h}{p - mv} \neq \lambda, \quad (1.7)$$

This is very strange! As we review now, for ordinary waves that propagate in the rest frame of a medium (like sound waves or water waves) Galilean observers will find frequency changes but no change in wavelength. This is intuitively clear: to find the wavelength one need only take a picture of the wave at some given time, and both observers looking at the picture will agree on the value of the wavelength. On the other hand to measure frequency, each observers must wait some time to see a full period of the wave go through them. This will take different time for the different observers.

Let us demonstrate these claims quantitatively. We begin with the statement that the phase $\phi = kx - \omega t$ of such a wave is a Galilean invariant. The wave itself may be $\cos \phi$ or $\sin \phi$ or some combination, but the fact is that the physical value of the wave at any point and time must be agreed by the two observers. The wave is an observable. Since all the features of the wave (peaks, zeroes, etc, etc) are controlled by the phase, the two observers must agree on the value of the phase.

In the S frame the phase can be written as follows

$$\phi = kx - \omega t = k(x - \frac{\omega}{k}t) = \frac{2\pi}{\lambda}(x - Vt) = \frac{2\pi x}{\lambda} - \frac{2\pi V}{\lambda}t, \quad (1.8)$$

where $V = \frac{\omega}{k}$ is the wave velocity. Note that the wavelength is read from the coefficient of x and ω is minus the coefficient of t . The two observers should agree on the value of ϕ . That is, we should have

$$\phi'(x', t') = \phi(x, t) \quad (1.9)$$

where the coordinates and times are related by a Galilean transformation. Therefore

$$\phi'(x', t') = \frac{2\pi}{\lambda}(x - Vt) = \frac{2\pi}{\lambda}(x' + vt' - Vt') = \frac{2\pi}{\lambda}x' - \frac{2\pi(V-v)}{\lambda}t'. \quad (1.10)$$

Since the right-hand side is expressed in terms of primed variables, we can read λ' from the coefficient of x' and ω' as minus the coefficient of t' :

$$\lambda' = \lambda \quad (1.11)$$

$$\omega' = \frac{2\pi}{\lambda}(V - v) = \frac{2\pi V}{\lambda} \left(1 - \frac{v}{V}\right) = \omega \left(1 - \frac{v}{V}\right). \quad (1.12)$$

This confirms that, as we claimed, for a physical wave propagating in a medium, the wavelength is a Galilean invariant and the frequency transforms.

So what does it mean that the wavelength of matter waves change under a Galilean transformation? It means that the Ψ waves are not directly measurable! Their value does not correspond to a measurable quantity for which all Galilean observers must agree. Thus, the wavefunction need not be invariant under Galilean transformations:

$$\Psi(x, t) \neq \Psi'(x', t'), \quad (1.13)$$

where (x, t) and (x', t') are related by Galilean transformations and thus represent the same point and time. You will figure out in Homework the correct relation between $\Psi(x, t)$ and $\Psi'(x', t')$.

What is the frequency ω of the de Broglie wave for a particle with momentum p ? We had

$$p = \hbar k \quad (1.14)$$

which fixes the wavelength in terms of the momentum. The frequency ω of the wave is determined by the relation

$$E = \hbar\omega, \quad (1.15)$$

which was also postulated by de Broglie and fixes ω in terms of the energy E of the particle. Note that for our focus on non-relativistic particles the energy E is determined by the momentum through the relation

$$E = \frac{p^2}{2m}. \quad (1.16)$$

We can give three pieces of evidence that (1.15) is a reasonable relation.

1. If we superpose matter waves to form a wave-packet that represents the particle, the packet will move with the so called group velocity v_g , which in fact coincides with the velocity of the particle. The group velocity is found by differentiation of ω with respect to k , as we will review soon:

$$v_g = \frac{d\omega}{dk} = \frac{dE}{dp} = \frac{d}{dp} \left(\frac{p^2}{2m} \right) = \frac{p}{m} = v. \quad (1.17)$$

2. The relation is also suggested by special relativity. The energy and the momentum components of a particle form a four-vector:

$$\left(\frac{E}{c}, p \right) \quad (1.18)$$

Similarly, for waves whose phases are relativistic invariant we have another four-vector

$$\left(\frac{\omega}{c}, k \right) \quad (1.19)$$

Setting two four-vectors equal to each other is a consistent choice: it would be valid in all Lorentz frames. As you can see, both de Broglie relations follow from

$$\left(\frac{E}{c}, p \right) = \hbar \left(\frac{\omega}{c}, k \right). \quad (1.20)$$

3. For photons, (1.15) is consistent with Einstein's quanta of energy, because $E = h\nu = \hbar\omega$.

In summary we have

$$p = \hbar k, \quad E = \hbar\omega. \quad (1.21)$$

These are called the *de Broglie relations*, and they are valid for all particles.

2 Phase and Group Velocities

To understand group velocity we form wave packets and investigate how fast they move. For this we will simply assume that $\omega(k)$ is some arbitrary function of k . Consider a superposition of plane waves $e^{i(kx-\omega(k)t)}$ given by

$$\psi(x, t) = \int dk \Phi(k) e^{i(kx-\omega(k)t)}. \quad (2.22)$$

We assume that the function $\Phi(k)$ is peaked around some wavenumber $k = k_0$, as shown in Fig. 2.

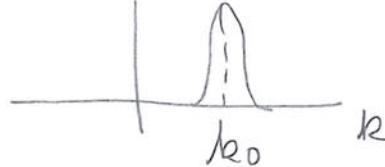


Figure 2: The function $\Phi(k)$ is assumed to peak around $k = k_0$.

In order to motivate the following discussion consider the case when $\Phi(k)$ not only peaks around k_0 but it also is *real* (we will drop this assumption later). In this case the phase φ of the integrand comes only from the exponential:

$$\varphi(k) = kx - \omega(k)t. \quad (2.23)$$

We wish to understand what are the values of x and t for which the packet $\psi(x, t)$ takes large values. We use the *stationary phase principle*: since only for $k \sim k_0$ the integral over k has a chance to give a non-zero contribution, the phase factor must be *stationary* at $k = k_0$. The idea is simple: if a function is multiplied by a rapidly varying phase, the integral washes out. Thus the phase must have zero derivative at k_0 . Applying this idea to our phase we find the derivative and set it equal to zero at k_0 :

$$\frac{d\varphi}{dk} \Big|_{k_0} = x - \frac{d\omega}{dk} \Big|_{k_0} t = 0. \quad (2.24)$$

This means that $\psi(x, t)$ is appreciable when x and t are related by

$$x = \frac{d\omega}{dk} \Big|_{k_0} t, \quad (2.25)$$

showing that the packet moves with *group velocity*

$$v_g = \frac{d\omega}{dk} \Big|_{k_0}. \quad (2.26)$$

Exercise. If $\Phi(k_0)$ is not real write $\Phi(k) = |\Phi(k)|e^{i\phi(k)}$. Find the new version of (2.25) and show that the velocity of the wave is not changed.

Let us now do a more detailed calculation that confirms the above analysis and gives some extra insight. Notice first that

$$\psi(x, 0) = \int dk \Phi(k) e^{ikx}. \quad (2.27)$$

We expand $\omega(k)$ in a Taylor expansion around $k = k_0$

$$\omega(k) = \omega(k_0) + (k - k_0) \frac{d\omega}{dk} \Big|_{k_0} + \mathcal{O}((k - k_0)^2). \quad (2.28)$$

Then we find, neglecting the $\mathcal{O}((k - k_0)^2)$ terms

$$\psi(x, t) = \int dk \Phi(k) e^{ikx} e^{-i\omega(k_0)t} e^{-i(k-k_0)\frac{d\omega}{dk}|_{k_0} t}. \quad (2.29)$$

It is convenient to take out of the integral all the factors that do not depend on k :

$$\begin{aligned} \psi(x, t) &= e^{-i\omega(k_0)t + ik_0 \frac{d\omega}{dk}|_{k_0} t} \int dk \Phi(k) e^{ikx} e^{-ik \frac{d\omega}{dk}|_{k_0} t} \\ &= e^{-i\omega(k_0)t + ik_0 \frac{d\omega}{dk}|_{k_0} t} \int dk \Phi(k) e^{ik(x - \frac{d\omega}{dk}|_{k_0} t)}. \end{aligned} \quad (2.30)$$

Comparing with (2.27) we realize that the integral in the above expression can be written in terms of the wavefunction at zero time:

$$\psi(x, t) = e^{-i\omega(k_0)t + ik_0 \frac{d\omega}{dk}|_{k_0} t} \psi\left(x - \frac{d\omega}{dk} \Big|_{k_0} t\right). \quad (2.31)$$

The phase factors in front of the expression are not important in tracking where the wave packet is. In particular we can take the norm of both sides of the equation to find

$$|\psi(x, t)| = \left| \psi\left(x - \frac{d\omega}{dk} \Big|_{k_0} t, 0\right) \right|. \quad (2.32)$$

If $\psi(x, 0)$ peaks at some value x_0 it is clear from the above equation that $|\psi(x, t)|$ peaks for

$$x - \frac{d\omega}{dk} \Big|_{k_0} t = x_0 \rightarrow x = x_0 + \frac{d\omega}{dk} \Big|_{k_0} t, \quad (2.33)$$

showing that the peak of the packet moves with velocity $v_{\text{gr}} = \frac{d\omega}{dk}$, evaluated at k_0 .

3 Choosing the wavefunction for a free particle

What is the mathematical form of the wave associated with a particle a particle with energy E and momentum p ? We know that ω and k are determined from $E = \hbar\omega$ and $p = \hbar k$. Let's suppose that we want our wave to be propagating in the $+\hat{x}$ direction. All the following are examples of waves that could be candidates for the particle wavefunction.

1. $\sin(kx - \omega t)$
2. $\cos(kx - \omega t)$
3. $e^{i(kx - \omega t)} = e^{ikx} e^{-i\omega t}$ - time dependence $\propto e^{-i\omega t}$
4. $e^{-i(kx - \omega t)} = e^{-ikx} e^{i\omega t}$ - time dependence $\propto e^{+i\omega t}$

In the third and fourth options we have indicated that the time dependence could come with either sign. We will use superposition to decide which is the right one! We are looking for a wave-function which is non-zero for all values of x .

Let's take them one by one:

1. Starting from (1), we build a superposition in which the particle has equal probability to be found moving in the $+x$ and the $-x$ directions.

$$\Psi(x, t) = \sin(kx - \omega t) + \sin(kx + \omega t) \quad (3.1)$$

Expanding the trigonometric functions this can be simplified to

$$\Psi(x, t) = 2 \sin(kx) \cos(\omega t). \quad (3.2)$$

But this result is not sensible. The wave function vanishes identically for all x at some special times

$$\omega t = \left(\frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \dots \right) \quad (3.3)$$

A wavefunction that is zero cannot represent a particle.

2. Constructing a wave function from (2) with a superposition of left and right going cos waves,

$$\Psi(x, t) = \cos(kx - \omega t) + \cos(kx + \omega t) = 2 \cos(kx) \cos(\omega t). \quad (3.4)$$

This choice is no good, it also vanishes identically when $\omega t = \left(\frac{\pi}{2}, \frac{3\pi}{2}, \dots \right)$

3. Let's try a similar superposition of exponentials from (3), with both having the same time dependence

$$\Psi(x, t) = e^{i(kx - \omega t)} + e^{i(-kx - \omega t)} \quad (3.5)$$

$$= (e^{ikx} + e^{-ikx}) e^{-i\omega t} \quad (3.6)$$

$$= 2 \cos(kx) e^{-i\omega t}. \quad (3.7)$$

This wavefunction meets our criteria! It is never zero for all values of x because $e^{-i\omega t}$ is never zero.

4. A superposition of exponentials from (4) also meets our criteria

$$\Psi(x, t) = e^{-i(kx - \omega t)} + e^{-i(-kx - \omega t)} \quad (3.8)$$

$$= (e^{ikx} + e^{-ikx}) e^{i\omega t} \quad (3.9)$$

$$= 2 \cos kx e^{i\omega t}. \quad (3.10)$$

This is never zero for all values of x

Since both options (3) and (4) seem to work we ask: Can we use *both* (3) and (4) to represent a particle moving to the right (in the $+\hat{x}$ direction)? Let's assume that we can. Then, since adding a state to itself should not change the state, we could represent the right moving particle by using the sum of (3) and (4)

$$\Psi(x, t) = e^{i(kx - \omega t)} + e^{-i(kx - \omega t)} = 2 \cos(kx - \omega t). \quad (3.11)$$

This, however, is the same as (2), which we already showed leads to difficulties. Therefore we must choose between (3) and (4).

The choice is a matter of convention, and all physicists use the same convention. We take the free particle wavefunction to be

$\text{Free particle wavefunction : } \Psi(x, t) = e^{i(kx - \omega t)},$

(3.12)

representing a particle with

$$p = \hbar k, \quad \text{and} \quad E = \hbar \omega. \quad (3.13)$$

In three dimensions the corresponding wavefunction would be

$\text{Free particle wavefunction : } \Psi(\mathbf{x}, t) = e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)},$

(3.14)

representing a particle with

$$\mathbf{p} = \hbar \mathbf{k}, \quad \text{and} \quad E = \hbar \omega. \quad (3.15)$$

Andrew Turner and Sarah Geller transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare
<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

Lecture 14 and 15: Algebraic approach to the SHO

B. Zwiebach
April 5, 2016

Contents

1 Algebraic Solution of the Oscillator	1
2 Operator manipulation and the spectrum	4

1 Algebraic Solution of the Oscillator

We have already seen how to calculate energy eigenstates for the simple harmonic oscillator by solving a second-order differential equation, the time-independent Schrödinger equation.

Let us now try to factorize the harmonic oscillator Hamiltonian. By this we mean, roughly, writing the Hamiltonian as the product of an operator times its *Hermitian conjugate*. As a first step we rewrite the Hamiltonian as

$$\hat{H} = \frac{1}{2}m\omega^2 \left(\hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2} \right). \quad (1.1)$$

Motivated by the identity $a^2 + b^2 = (a - ib)(a + ib)$, holding for numbers a and b , we examine if the expression in parenthesis can be written as a product

$$\begin{aligned} \left(\hat{x} - \frac{i\hat{p}}{m\omega} \right) \left(\hat{x} + \frac{i\hat{p}}{m\omega} \right) &= \hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2} + \frac{i}{m\omega} (\hat{x}\hat{p} - \hat{p}\hat{x}), \\ &= \hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2} - \frac{\hbar}{m\omega} \mathbf{1}, \end{aligned} \quad (1.2)$$

where the extra terms arise because \hat{x} and \hat{p} , as opposed to numbers, do not commute. We now define the right-most factor in the above product to be V :

$$V \equiv \hat{x} + \frac{i\hat{p}}{m\omega}, \quad (1.3)$$

Since \hat{x} and \hat{p} are Hermitian operators, we then have

$$V^\dagger = \hat{x} - \frac{i\hat{p}}{m\omega}, \quad (1.4)$$

and this is the left-most factor in the product! We can therefore rewrite (1.2) as

$$\hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2} = V^\dagger V + \frac{\hbar}{m\omega} \mathbf{1}, \quad (1.5)$$

and therefore back in the Hamiltonian (1.1) we find,

$$\hat{H} = \frac{1}{2}m\omega^2 V^\dagger V + \frac{1}{2}\hbar\omega \mathbf{1}. \quad (1.6)$$

This is a factorized form of the Hamiltonian: up to an additive constant E_0 , \hat{H} is the product of a positive constant times the operator product $V^\dagger V$. We note that the commutator of V and V^\dagger is simple

$$[V, V^\dagger] = \left[\hat{x} + \frac{i\hat{p}}{m\omega}, \hat{x} - \frac{i\hat{p}}{m\omega} \right] = -\frac{i}{m\omega}[\hat{x}, \hat{p}] + \frac{i}{m\omega}[\hat{p}, \hat{x}] = \frac{2\hbar}{m\omega}\mathbf{1}. \quad (1.7)$$

This implies that

$$\left[\sqrt{\frac{m\omega}{2\hbar}}V, \sqrt{\frac{m\omega}{2\hbar}}V^\dagger \right] = \mathbf{1}. \quad (1.8)$$

This suggests the definition of unit-free operator operators \hat{a} and \hat{a}^\dagger :

$$\begin{aligned} \hat{a} &\equiv \sqrt{\frac{m\omega}{2\hbar}}V, \\ \hat{a}^\dagger &\equiv \sqrt{\frac{m\omega}{2\hbar}}V^\dagger. \end{aligned} \quad (1.9)$$

Due to the scaling we have

$$[\hat{a}, \hat{a}^\dagger] = \mathbf{1}. \quad (1.10)$$

The operator \hat{a} is called *annihilation* operator and \hat{a}^\dagger is called a *creation* operator. The justification for these names will be seen below. From the above definitions we read the relations between $(\hat{a}, \hat{a}^\dagger)$ and (\hat{x}, \hat{p}) :

$$\boxed{\begin{aligned} \hat{a} &= \sqrt{\frac{m\omega}{2\hbar}} \left(\hat{x} + \frac{i\hat{p}}{m\omega} \right), \\ \hat{a}^\dagger &= \sqrt{\frac{m\omega}{2\hbar}} \left(\hat{x} - \frac{i\hat{p}}{m\omega} \right). \end{aligned}} \quad (1.11)$$

The inverse relations are many times useful as well,

$$\boxed{\begin{aligned} \hat{x} &= \sqrt{\frac{\hbar}{2m\omega}}(\hat{a} + \hat{a}^\dagger), \\ \hat{p} &= i\sqrt{\frac{m\omega\hbar}{2}}(\hat{a}^\dagger - \hat{a}). \end{aligned}} \quad (1.12)$$

While neither \hat{a} nor \hat{a}^\dagger is hermitian (they are hermitian conjugates of each other), the above equations are consistent with the hermiticity of \hat{x} and \hat{p} . We can now write the Hamiltonian in terms of the \hat{a} and \hat{a}^\dagger operators. Using (1.9) we have

$$V^\dagger V = \frac{2\hbar}{m\omega} \hat{a}^\dagger \hat{a}, \quad (1.13)$$

and therefore back in (1.6) we get

$$\boxed{\hat{H} = \hbar\omega(\hat{a}^\dagger \hat{a} + \frac{1}{2}) = \hbar\omega(\hat{N} + \frac{1}{2}), \quad \hat{N} \equiv \hat{a}^\dagger \hat{a}.} \quad (1.14)$$

The above form of the Hamiltonian is factorized: up to an additive constant \hat{H} is the product of a positive constant times the operator product $\hat{a}^\dagger \hat{a}$. In here we have dropped the identity operator, which

is usually understood. We have also introduced the *number* operator \hat{N} . This is, by construction, a hermitian operator and it is, up to a scale and an additive constant, equal to the Hamiltonian. An eigenstate of \hat{H} is also an eigenstate of \hat{N} and it follows from the above relation that the respective eigenvalues E and N are related by

$$E = \hbar\omega \left(N + \frac{1}{2} \right). \quad (1.15)$$

Let us now show the powerful conclusions that arise from the factorized Hamiltonian. On any state ψ that is normalized we have

$$\langle \hat{H} \rangle_\psi = (\psi, \hat{H}\psi) = \hbar\omega(\psi, \hat{a}^\dagger \hat{a}\psi) + \frac{1}{2}\hbar\omega(\psi, \psi), \quad (1.16)$$

and moving the \hat{a}^\dagger to the first input, we get

$$\langle \hat{H} \rangle_\psi = \hbar\omega(\hat{a}\psi, \hat{a}\psi) + \frac{1}{2}\hbar\omega \geq \frac{1}{2}\hbar\omega. \quad (1.17)$$

The inequality follows because any expression of the form (φ, φ) is greater than or equal to zero. This shows that for any energy eigenstate with energy E : $\hat{H}\psi = E\psi$ we have

$$\text{Energy eigenstates: } E \geq \frac{1}{2}\hbar\omega. \quad (1.18)$$

This important result about the spectrum followed directly from the factorization of the Hamiltonian. But we also get the information required to find the ground state wavefunction. The minimum energy $\frac{1}{2}\hbar\omega$ will be realized for a state ψ if the term $(\hat{a}\psi, \hat{a}\psi)$ in (1.17) vanishes. For this to vanish $\hat{a}\psi$ must vanish. Therefore, the ground state wavefunction φ_0 must satisfy

$$\hat{a}\varphi_0 = 0. \quad (1.19)$$

The operator \hat{a} annihilates the ground state and this why \hat{a} is called the annihilation operator. Using the definition of \hat{a} in (1.11) and the position space representation of \hat{p} , this becomes

$$\left(x + \frac{i}{m\omega} \frac{\hbar}{i} \frac{d}{dx} \right) \varphi_0(x) = 0 \rightarrow \left(x + \frac{\hbar}{m\omega} \frac{d}{dx} \right) \varphi_0(x) = 0. \quad (1.20)$$

Remarkably, this is a **first order** differential equation for the ground state. Not a second order equation, like the Schrödinger equation that determines the general energy eigenstates. This is a dramatic simplification afforded by the factorization of the Hamiltonian into a product of first-order differential operators. The above equation is rearranged as

$$\frac{d\varphi_0}{dx} = -\frac{m\omega}{\hbar} x \varphi_0. \quad (1.21)$$

Solving this differential equation yields

$$\varphi_0(x) = \left(\frac{m\omega}{\pi\hbar} \right)^{\frac{1}{4}} e^{-\frac{m\omega}{2\hbar}x^2},$$

(1.22)

where we included a normalization constant to guarantee that $(\varphi_0, \varphi_0) = 1$. Note that φ_0 is indeed an energy eigenstate with energy E_0 :

$$\hat{H}\varphi_0 = \hbar\omega(\hat{a}^\dagger \hat{a} + \frac{1}{2})\varphi_0 = \frac{1}{2}\hbar\omega\varphi_0 \rightarrow E_0 = \frac{1}{2}\hbar\omega. \quad (1.23)$$

Before proceeding with the analysis of excited states, let us view the properties of factorization more generally. Factorizing a Hamiltonian means finding an operator \hat{A} such that we can rewrite the Hamiltonian as $\hat{A}^\dagger \hat{A}$ up to an additive constant. Here \hat{A}^\dagger is the Hermitian conjugate of \hat{A} , an operator that is defined by

$$(\psi, \hat{A}^\dagger \varphi) = (\hat{A}\psi, \varphi). \quad (1.24)$$

We say that we have factorized a Hamiltonian \hat{H} if we can find a \hat{A} for which

$$\boxed{\hat{H} = \hat{A}^\dagger \hat{A} + E_0 \mathbf{1}}, \quad (1.25)$$

where E_0 is a constant with units of energy that multiplies the identity operator. This constant does not complicate our task of finding the eigenstates of the Hamiltonian, nor their energies: any eigenfunction of $\hat{A}^\dagger \hat{A}$ is an eigenfunction of \hat{H} . Two key properties follow from the factorization (1.25).

1. Any energy eigenstate must have energy greater than or equal to E_0 . First note that for an *arbitrary* normalized $\psi(x)$ we have

$$(\psi, \hat{H}\psi) = (\psi, \hat{A}^\dagger \hat{A}\psi) + E_0(\psi, \psi) = (\hat{A}\psi, \hat{A}\psi) + E_0, \quad (1.26)$$

Since the overlap $(\hat{A}\psi, \hat{A}\psi)$ is greater than or equal to zero, we have shown that

$$\boxed{(\psi, \hat{H}\psi) \geq E_0}. \quad (1.27)$$

If we take ψ to be an energy eigenstate of energy E : $\hat{H}\psi = E\psi$, the above relation gives

$$E \geq E_0. \quad (1.28)$$

This shows, as claimed, that all possible energies are greater than or equal to E_0 .

2. A wavefunction ψ_0 that satisfies

$$\hat{A}\psi_0 = 0, \quad (1.29)$$

is an energy eigenstate that *saturates* the inequality (1.28). Indeed,

$$\hat{H}\psi_0 = \hat{A}^\dagger \hat{A}\psi_0 + E_0\psi_0 = \hat{A}^\dagger(\hat{A}\psi_0) + E_0\psi_0 = E_0\psi_0. \quad (1.30)$$

The state ψ_0 satisfying $\hat{A}\psi_0 = 0$ is the ground state. For conventional Hamiltonians this is a first order differential equation for ψ_0 and much easier to solve than the Schrödinger equation.

2 Operator manipulation and the spectrum

We have seen that all energy eigenstates are eigenstates of the Hermitian number operator $\hat{N} = \hat{a}^\dagger \hat{a}$. This is because $\hat{H} = \hbar\omega(\hat{N} + \frac{1}{2})$. Note that since $\hat{a}\varphi_0 = 0$ we also have

$$\hat{N}\varphi_0 = 0. \quad (2.1)$$

We can quickly check that

$$\begin{aligned} [\hat{N}, \hat{a}] &= [\hat{a}^\dagger \hat{a}, \hat{a}] = [\hat{a}^\dagger, \hat{a}] \hat{a} = -\hat{a}, \\ [\hat{N}, \hat{a}^\dagger] &= [\hat{a}^\dagger \hat{a}, \hat{a}^\dagger] = \hat{a}^\dagger [\hat{a}, \hat{a}^\dagger] = \hat{a}^\dagger, \end{aligned} \quad (2.2)$$

which we summarize as

$$\begin{aligned} [\hat{N}, \hat{a}] &= -\hat{a}, \\ [\hat{N}, \hat{a}^\dagger] &= \hat{a}^\dagger. \end{aligned} \quad (2.3)$$

Using these identities and induction you should be able to show that:

$$\begin{aligned} [\hat{N}, (\hat{a})^k] &= -k(\hat{a})^k, \\ [\hat{N}, (\hat{a}^\dagger)^k] &= k(\hat{a}^\dagger)^k. \end{aligned} \quad (2.4)$$

These relations suggest why \hat{N} is called the number operator. Acting on powers of creation or annihilation operators by commutation it gives the same object multiplied by (plus or minus) the number of creation or annihilation operators, k in the above. Closely related commutators are also useful:

$$\begin{aligned} [\hat{a}^\dagger, (\hat{a})^k] &= -k(\hat{a})^{k-1} \\ [\hat{a}, (\hat{a}^\dagger)^k] &= k(\hat{a}^\dagger)^{k-1}. \end{aligned} \quad (2.5)$$

These commutators are analogous to $[\hat{p}, (\hat{x})^k]$ and $[\hat{x}, (\hat{p})^k]$. We will also make use of the following Lemma which helps in evaluations where we have an operator \hat{A} that kills a state ψ and we aim to simplify the action of $\hat{A}\hat{B}$, where \hat{B} is another operator, acting on ψ . Here is the result

If $\hat{A}\psi = 0$, then $\hat{A}\hat{B}\psi = [\hat{A}, \hat{B}]\psi$.

(2.6)

This is easily proved. First note that

$$\hat{A}\hat{B} = [\hat{A}, \hat{B}] + \hat{B}\hat{A}, \quad (2.7)$$

as can be quickly checked expanding the right-hand side. It then follows that

$$\hat{A}\hat{B}\psi = ([\hat{A}, \hat{B}] + \hat{B}\hat{A})\psi = [\hat{A}, \hat{B}]\psi, \quad (2.8)$$

because $\hat{B}\hat{A}\psi = \hat{B}(\hat{A}\psi) = 0$. This is what we wanted to show. This is all we need to know about commutators and we can now proceed to construct the states of the harmonic oscillator.

Since \hat{a} annihilates φ_0 consider acting on the ground state with \hat{a}^\dagger . It is clear that \hat{a}^\dagger cannot also annihilate φ_0 . If that would happen acting with both sides of the commutator identity $[\hat{a}, \hat{a}^\dagger] = 1$ on φ_0 would lead to a contradiction: the left-hand side would vanish but the right-hand side would not. Thus consider the wavefunction

$$\varphi_1 \equiv \hat{a}^\dagger \varphi_0. \quad (2.9)$$

We are going to show that this is an energy eigenstate. For this purpose we act on it with the number operator:

$$\hat{N}\varphi_1 = \hat{N}\hat{a}^\dagger \varphi_0 = [\hat{N}, \hat{a}^\dagger]\varphi_0, \quad (2.10)$$

where we noted that $\hat{N}\varphi_0 = 0$ and used Lemma (2.6). Given that $[\hat{N}, \hat{a}^\dagger] = \hat{a}^\dagger$, we get

$$\hat{N}\varphi_1 = \hat{a}^\dagger\varphi_0 = \varphi_1. \quad (2.11)$$

Thus φ_1 is an eigenstate of the operator \hat{N} with eigenvalue $N = 1$. Since φ_0 has \hat{N} eigenvalue zero, the effect of acting on φ_0 with \hat{a}^\dagger was to increase the eigenvalue of the number operator by one unit. The operator \hat{a}^\dagger is called the *creation* operator because it creates a state out of the ground state. Alternatively, it is called the *raising* operator, because it raises (by one unit) the eigenvalue of \hat{N} . Since $N = 1$ for φ_1 it follows that φ_1 is an energy eigenstate with energy E_1 given by

$$E_1 = \hbar\omega(1 + \frac{1}{2}) = \frac{3}{2}\hbar\omega. \quad (2.12)$$

It also turns out that φ_1 is properly normalized:

$$(\varphi_1, \varphi_1) = (\hat{a}^\dagger\varphi_0, \hat{a}^\dagger\varphi_0) = (\varphi_0, \hat{a}\hat{a}^\dagger\varphi_0), \quad (2.13)$$

where we used the Hermitian conjugation property to move the \hat{a}^\dagger acting on the left input into the right input, where it goes as $(\hat{a}^\dagger)^\dagger = \hat{a}$. We then have

$$(\varphi_1, \varphi_1) = (\varphi_0, \hat{a}\hat{a}^\dagger\varphi_0) = (\varphi_0, [\hat{a}, \hat{a}^\dagger]\varphi_0) = (\varphi_0, \varphi_0) = 1, \quad (2.14)$$

where we used (2.6) in the evaluation of $\hat{a}\hat{a}^\dagger\psi_0$. Indeed the state φ_1 is correctly normalized.

Next consider the state

$$\varphi'_2 \equiv \hat{a}^\dagger\hat{a}^\dagger\varphi_0. \quad (2.15)$$

This has

$$\hat{N}\varphi'_2 = \hat{N}\hat{a}^\dagger\hat{a}^\dagger\varphi_0 = [\hat{N}, \hat{a}^\dagger\hat{a}^\dagger]\varphi_0 = 2\hat{a}^\dagger\hat{a}^\dagger\varphi_0 = 2\varphi'_2, \quad (2.16)$$

so φ_2 is a state with number $N = 2$ and energy $E_2 = \frac{5}{2}\hbar\omega$. Is it properly normalized? We find

$$\begin{aligned} (\varphi_2, \varphi_2) &= (\hat{a}^\dagger\hat{a}^\dagger\varphi_0, \hat{a}^\dagger\hat{a}^\dagger\varphi_0) = (\varphi_0, \hat{a}\hat{a}\hat{a}^\dagger\hat{a}^\dagger\varphi_0) = (\varphi_0, \hat{a}[\hat{a}, \hat{a}^\dagger]\hat{a}^\dagger\varphi_0) \\ &= (\varphi_0, 2\hat{a}\hat{a}^\dagger\varphi_0) = (\varphi_0, \varphi_0) = 2. \end{aligned} \quad (2.17)$$

The properly normalized wavefunction is therefore

$$\varphi_2 \equiv \frac{1}{\sqrt{2}}\hat{a}^\dagger\hat{a}^\dagger\varphi_0. \quad (2.18)$$

We now claim that the n -th excited state of the simple harmonic oscillator is

$$\varphi_n \equiv \frac{1}{\sqrt{n!}} \underbrace{\hat{a}^\dagger \cdots \hat{a}^\dagger}_n \varphi_0 = \frac{1}{\sqrt{n!}} (\hat{a}^\dagger)^n \varphi_0. \quad (2.19)$$

Exercise: Verify that this state has \hat{N} eigenvalue n .

Exercise: Verify that the state φ_n is properly normalized.

Since the \hat{N} eigenvalue of φ is n , its energy E_n is given by

$$E_n = \hbar\omega(n + \frac{1}{2}). \quad (2.20)$$

Since the various states φ_n are eigenstates of a Hermitian operator (the Hamiltonian \hat{H}) with different eigenvalues, they are orthonormal

$$(\varphi_n, \varphi_m) = \delta_{m,n}. \quad (2.21)$$

We now note that $\hat{a}\varphi_n$ is a state with $n - 1$ operators \hat{a}^\dagger acting on φ_0 because the \hat{a} eliminates one of the creation operators in φ_n . Thus we expect $\hat{a}\varphi_n \sim \varphi_{n-1}$. We can make this precise

$$\hat{a}\varphi_n = \hat{a}\frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n\varphi_0 = \frac{1}{\sqrt{n!}}[\hat{a}, (\hat{a}^\dagger)^n]\varphi_0 = \frac{n}{\sqrt{n!}}(\hat{a}^\dagger)^{n-1}\varphi_0. \quad (2.22)$$

At this point we use (2.19) with n set equal to $n - 1$ and thus get

$$\hat{a}\varphi_n = \frac{n}{\sqrt{n!}}\sqrt{(n-1)!}\varphi_{n-1} = \sqrt{n}\varphi_{n-1}. \quad (2.23)$$

By the action of \hat{a}^\dagger on φ_n we get

$$\hat{a}^\dagger\varphi_n = \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^{n+1}\varphi_0 = \frac{1}{\sqrt{n!}}\sqrt{(n+1)!}\varphi_{n+1} = \sqrt{n+1}\varphi_{n+1}. \quad (2.24)$$

Collecting the results, we have

$$\begin{aligned} \hat{a}\varphi_n &= \sqrt{n}\varphi_{n-1}, \\ \hat{a}^\dagger\varphi_n &= \sqrt{n+1}\varphi_{n+1}. \end{aligned}$$

(2.25)

These relations make it clear that \hat{a} lowers the number of any energy eigenstate by one unit, except for the vacuum φ_0 which it kills. The raising operator \hat{a}^\dagger increases the number of any eigenstate by one unit.

Exercise: Calculate the uncertainty Δx of position in the n -th energy eigenstate.

Solution: By definition,

$$(\Delta x)_n^2 = \langle \hat{x}^2 \rangle_{\varphi_n} - \langle \hat{x} \rangle_{\varphi_n}^2. \quad (2.26)$$

The expectation value $\langle \hat{x} \rangle$ vanishes for any energy eigenstate since we are integrating x , which is odd, against $|\varphi_n(x)|^2$, which is always even. Still, it is instructive to see how this happens explicitly:

$$\langle \hat{x} \rangle_{\varphi_n} = (\varphi_n, \hat{x}\varphi_n) = \sqrt{\frac{\hbar}{2m\omega}}(\varphi_n, (\hat{a} + \hat{a}^\dagger)\varphi_n), \quad (2.27)$$

using the formula for \hat{x} in terms of \hat{a} and \hat{a}^\dagger . The above overlap vanishes because $\hat{a}\varphi_n \sim \varphi_{n-1}$ and $\hat{a}^\dagger\varphi_n \sim \varphi_{n+1}$ and both φ_{n-1} and φ_{n+1} are orthogonal to φ_n . Now we compute the expectation value of \hat{x}^2

$$\begin{aligned} \langle \hat{x}^2 \rangle_{\varphi_n} &= (\varphi_n, \hat{x}^2\varphi_n) = \frac{\hbar}{2m\omega}(\varphi_n, (\hat{a} + \hat{a}^\dagger)(\hat{a} + \hat{a}^\dagger)\varphi_n) \\ &= \frac{\hbar}{2m\omega}(\varphi_n, (\hat{a}\hat{a} + \hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a} + \hat{a}^\dagger\hat{a}^\dagger)\varphi_n). \end{aligned} \quad (2.28)$$

Since $\hat{a}\hat{a}\varphi_n \sim \varphi_{n-2}$ and $\hat{a}^\dagger\hat{a}^\dagger\varphi_n \sim \varphi_{n+2}$ and both φ_{n-2} and φ_{n+2} are orthogonal to φ_n , the $\hat{a}\hat{a}$ and $\hat{a}^\dagger\hat{a}^\dagger$ terms do not contribute. We are left with

$$\langle \hat{x}^2 \rangle_{\varphi_n} = \frac{\hbar}{2m\omega}(\varphi_n, (\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a})\varphi_n). \quad (2.29)$$

At this point we recognize that $\hat{a}^\dagger \hat{a} = \hat{N}$ and that $\hat{a}\hat{a}^\dagger = [\hat{a}, \hat{a}^\dagger] + \hat{a}^\dagger \hat{a} = 1 + \hat{N}$. As a result

$$\langle \hat{x}^2 \rangle_{\varphi_n} = \frac{\hbar}{2m\omega} (\varphi_n, (1 + 2\hat{N})\varphi_n) = \frac{\hbar}{2m\omega} (1 + 2n). \quad (2.30)$$

We therefore have

$$(\Delta x)_n^2 = \frac{\hbar}{m\omega} \left(n + \frac{1}{2} \right). \quad (2.31)$$

The position uncertainty grows linearly with the number.

Sarah Geller and Andrew Turner transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare

<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

DIRAC's BRA AND KET NOTATION

B. Zwiebach

October 7, 2013

Contents

1 From inner products to bra-kets	1
2 Operators revisited	5
2.1 Projection Operators	6
2.2 Adjoint of a linear operator	8
2.3 Hermitian and Unitary Operators	9
3 Non-denumerable basis	11

1 From inner products to bra-kets

Dirac invented a useful alternative notation for inner products that leads to the concepts of bras and kets. The notation is sometimes more efficient than the conventional mathematical notation we have been using. It is also widely although not universally used. It all begins by writing the inner product differently. The rule is to turn inner products into bra-ket pairs as follows

$$\langle u, v \rangle \longrightarrow \langle u | v \rangle. \quad (1.1)$$

Instead of the inner product comma we simply put a vertical bar! We can translate our earlier discussion of inner products trivially. In order to make you familiar with the new look we do it.

We now write $\langle u | v \rangle = \langle v | u \rangle^*$, as well as $\langle v | v \rangle \geq 0$ for all v , while $\langle v | v \rangle = 0$ if and only if $v = 0$. We have linearity in the second argument

$$\langle u | c_1 v_1 + c_2 v_2 \rangle = c_1 \langle u | v_1 \rangle + c_2 \langle u | v_2 \rangle, \quad (1.2)$$

for complex constants c_1 and c_2 , but antilinearity in the first argument

$$\langle c_1 u_1 + c_2 u_2 | v \rangle = c_1^* \langle u_1 | v \rangle + c_2^* \langle u_2 | v \rangle. \quad (1.3)$$

Two vectors u and v for which $\langle u | v \rangle = 0$ are orthogonal. For the norm: $|v|^2 = \langle v | v \rangle$. The Schwarz inequality, for any pair u and v of vectors reads $|\langle u | v \rangle| \leq |u| |v|$.

For a given physical situation, the inner product must be defined and should satisfy the axioms. Let us consider two examples:

1. Let $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ and $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ be two vectors in a complex dimensional vector space of dimension two. We then define

$$\langle a|b\rangle \equiv a_1^*b_1 + a_2^*b_2. \quad (1.4)$$

You should confirm the axioms are satisfied.

2. Consider the complex vector space of complex function $f(x) \in \mathbb{C}$ with $x \in [0, L]$. Given two such functions $f(x), g(x)$ we define

$$\langle f|g\rangle \equiv \int_0^L f^*(x)g(x)dx. \quad (1.5)$$

The verification of the axioms is again quite straightforward.

A set of basis vectors $\{e_i\}$ labelled by the integers $i = 1, \dots, n$ satisfying

$$\langle e_i|e_j\rangle = \delta_{ij}, \quad (1.6)$$

is orthonormal. An arbitrary vector can be written as a linear superposition of basis states:

$$v = \sum_i \alpha_i e_i, \quad (1.7)$$

We then see that the coefficients are determined by the inner product

$$\langle e_k|v\rangle = \langle e_k|\sum_i \alpha_i e_i\rangle = \sum_i \alpha_i \langle e_k|e_i\rangle = \alpha_k. \quad (1.8)$$

We can therefore write

$$v = \sum_i e_i \langle e_i|v\rangle. \quad (1.9)$$

To obtain now bras and kets, we reinterpret the inner product. We want to “split” the inner product into two ingredients

$$\langle u|v\rangle \rightarrow \langle u| |v\rangle. \quad (1.10)$$

Here $|v\rangle$ is called a **ket** and $\langle u|$ is called a **bra**. We will view the ket $|v\rangle$ just as another way to represent the vector v . This is a small subtlety with the notation: we think of $v \in V$ as a vector and also $|v\rangle \in V$ as a vector. It is as if we added some decoration $| \rangle$ around the vector v to make it clear by inspection that it is a vector, perhaps like the usual top arrows that are added in some cases. The label in the ket is a vector and the ket itself is that vector!

Bras are somewhat different objects. We say that bras belong to the space V^* dual to V . Elements of V^* are linear maps from V to \mathbb{C} . In conventional mathematical notation one has a $v \in V$ and a linear function $\phi \in V^*$ such that $\phi(v)$, which denotes the action of the function of the vector v , is a number. In the bracket notation we have the replacements

$$\begin{aligned} v &\rightarrow |v\rangle, \\ \phi &\rightarrow \langle u|, \\ \phi_u(v) &\rightarrow \langle u|v\rangle, \end{aligned} \quad (1.11)$$

where we used the notation in (6.6). Our bras are labelled by vectors: the object inside the $\langle |$ is a vector. But bras are *not* vectors. If kets are viewed as column vectors, then bras are viewed as row vectors. In this way a bra to the left of a ket makes sense: matrix multiplication of a row vector times a column vector gives a number. Indeed, for vectors

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \quad (1.12)$$

we had

$$\langle a|b\rangle = a_1^*b_1 + a_2^*b_2 + \dots + a_n^*b_n \quad (1.13)$$

Now we think of this as

$$\langle a| = (a_1^*, a_2^*, \dots, a_n^*), \quad |b\rangle = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \quad (1.14)$$

and matrix multiplication gives us the desired answer

$$\langle a|b\rangle = (a_1^*, a_2^*, \dots, a_n^*) \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = a_1^*b_1 + a_2^*b_2 + \dots + a_n^*b_n. \quad (1.15)$$

Note that the bra labeled by the vector a is obtained by forming the row vector and complex conjugating the entries. More abstractly the bra $\langle u|$ labeled by the vector u is defined by its action on arbitrary vectors $|v\rangle$ as follows

$$\langle u| : |v\rangle \rightarrow \langle u|v\rangle. \quad (1.16)$$

As required by the definition, any linear map from V to \mathbb{C} defines a bra, and the corresponding underlying vector. For example let v be a generic vector:

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, \quad (1.17)$$

A linear map $f(v)$ that acting on a vector v gives a number is an expression of the form

$$f(v) = \alpha_1^*v_1 + \alpha_2^*v_2 + \dots + \alpha_n^*v_n. \quad (1.18)$$

It is a linear function of the components of the vector. The linear function is specified by the numbers α_i , and for convenience (and without loss of generality) we used their complex conjugates. Note that we need exactly n constants, so they can be used to assemble a row vector or a bra

$$\langle \alpha| = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*) \quad (1.19)$$

and the associated vector or ket

$$|\alpha\rangle = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \quad (1.20)$$

Note that, by construction

$$f(v) = \langle \alpha | v \rangle. \quad (1.21)$$

This illustrates the point that (i) bras represent dual objects that act on vectors and (ii) bras are labelled by vectors.

Bras can be added and can be multiplied by complex numbers and there is a zero bra defined to give zero acting on any vector, so V^* is also a complex vector space. As a bra, the linear superposition

$$\langle \omega | \equiv \alpha \langle a | + \beta \langle b | \in V^*, \quad \alpha, \beta \in \mathbb{C}, \quad (1.22)$$

is defined to act on a vector (ket) $|c\rangle$ to give the number

$$\alpha \langle a | c \rangle + \beta \langle b | c \rangle. \quad (1.23)$$

For any vector $|v\rangle \in V$ there is a *unique* bra $\langle v | \in V^*$. If there would be another bra $\langle v' |$ it would have to act on arbitrary vectors $|w\rangle$ just like $\langle v |$:

$$\langle v' | w \rangle = \langle v | w \rangle \rightarrow \langle w | v \rangle - \langle w | v' \rangle = 0 \rightarrow \langle w | v - v' \rangle = 0. \quad (1.24)$$

In the first step we used complex conjugation and in the second step linearity. Now the vector $v - v'$ must have zero inner product with *any* vector w , so $v - v' = 0$ and $v = v'$.

We can now reconsider equation (1.3) and write an extra right-hand side

$$\langle \alpha_1 a_1 + \alpha_2 a_2 | b \rangle = \alpha_1^* \langle a_1 | b \rangle + \alpha_2^* \langle a_2 | b \rangle = (\alpha_1^* \langle a_1 | + \alpha_2^* \langle a_2 |) |b\rangle \quad (1.25)$$

so that we conclude that the rules to pass from kets to bras include

$$|v\rangle = \alpha_1 |a_1\rangle + \alpha_2 |a_2\rangle \longleftrightarrow \langle v | = \alpha_1^* \langle a_1 | + \alpha_2^* \langle a_2 |. \quad (1.26)$$

For simplicity of notation we sometimes write kets with labels simpler than vectors. Let us reconsider the basis vectors $\{e_i\}$ discussed in (1.6). The ket $|e_i\rangle$ is simply called $|i\rangle$ and the orthonormal condition reads

$$\langle i | j \rangle = \delta_{ij}. \quad (1.27)$$

The expansion (1.7) of a vector now reads

$$|v\rangle = \sum_i |i\rangle \alpha_i, \quad (1.28)$$

As in (1.8) the expansion coefficients are $\alpha_k = \langle k | v \rangle$ so that

$$|v\rangle = \sum_i |i\rangle \langle i | v \rangle. \quad (1.29)$$

2 Operators revisited

Let T be an operator in a vector space V . This means that acting on vectors on V it gives vectors on V , something we write as

$$\Omega : V \rightarrow V. \quad (2.30)$$

We denote by $\Omega|a\rangle$ the vector obtained by acting with Ω on the vector $|a\rangle$:

$$|a\rangle \in V \rightarrow \Omega|a\rangle \in V. \quad (2.31)$$

The operator Ω is linear if additionally we have

$$\Omega(|a\rangle + |b\rangle) = \Omega|a\rangle + \Omega|b\rangle, \text{ and } \Omega(\alpha|a\rangle) = \alpha\Omega|a\rangle. \quad (2.32)$$

When kets are labeled by vectors we sometimes write

$$|\Omega a\rangle \equiv \Omega|a\rangle, \quad (2.33)$$

It is useful to note that a linear operator on V is also a linear operator on V^*

$$\Omega : V^* \rightarrow V^*, \quad (2.34)$$

We write this as

$$\langle a| \rightarrow \langle a|\Omega \in V^*. \quad (2.35)$$

The object $\langle a|\Omega$ is defined to be the bra that acting on the ket $|b\rangle$ gives the number $\langle a|\Omega|b\rangle$.

We can write operators in terms of bras and kets, written in a suitable order. As an example of an operator consider a bra $\langle a|$ and a ket $|b\rangle$. We claim that the object

$$\Omega = |a\rangle\langle b|, \quad (2.36)$$

is naturally viewed as a linear operator on V and on V^* . Indeed, acting on a vector we let it act as the bra-ket notation suggests:

$$\Omega|v\rangle \equiv |a\rangle\langle b|v\rangle \sim |a\rangle, \text{ since } \langle b|v\rangle \text{ is a number.} \quad (2.37)$$

Acting on a bra it gives a bra:

$$\langle w|\Omega \equiv \langle w|a\rangle\langle b| \sim \langle b|, \text{ since } \langle w|a\rangle \text{ is a number.} \quad (2.38)$$

Let us now review the description of operators as matrices. The choice of basis is ours to make. For simplicity, however, we will usually consider orthonormal bases.

Consider therefore, two vectors expanded in an orthonormal basis $\{|i\rangle\}$:

$$|a\rangle = \sum_n |n\rangle a_n, \quad |b\rangle = \sum_n |n\rangle b_n. \quad (2.39)$$

Assume $|b\rangle$ is obtained by the action of Ω on $|a\rangle$:

$$\Omega|a\rangle = |b\rangle \rightarrow \sum_n \Omega|n\rangle a_n = \sum_n |n\rangle b_n. \quad (2.40)$$

Acting on both sides of this vector equation with the bra $\langle m|$ we find

$$\sum_n \langle m|\Omega|n\rangle a_n = \sum_n \langle m|n\rangle b_n = b_m \quad (2.41)$$

We now define the ‘matrix elements’

$$\boxed{\Omega_{mn} \equiv \langle m|\Omega|n\rangle.} \quad (2.42)$$

so that the above equation reads

$$\sum_n \Omega_{mn} a_n = b_m, \quad (2.43)$$

which is the matrix version of the original relation $\Omega|a\rangle = |b\rangle$. The chosen basis has allowed us to view the linear operator Ω as a matrix, also denoted as Ω , with matrix components Ω_{mn} :

$$\Omega \longleftrightarrow \begin{pmatrix} \Omega_{11} & \Omega_{12} & \dots & \dots & \Omega_{1N} \\ \Omega_{21} & \Omega_{22} & \dots & \dots & \Omega_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Omega_{N1} & \Omega_{N2} & \dots & \dots & \Omega_{NN} \end{pmatrix}, \quad \text{with } \Omega_{ij} = \langle i|\Omega|j\rangle. \quad (2.44)$$

There is one additional claim. The operator itself can be written in terms of the matrix elements and basis bras and kets. We claim that

$$\boxed{\Omega = \sum_{m,n} |m\rangle \Omega_{mn} \langle n|.} \quad (2.45)$$

We can verify that this is correct by computing the matrix elements using it:

$$\langle m'|\Omega|n'\rangle = \sum_{m,n} \Omega_{mn} \langle m'|m\rangle \langle n|n'\rangle = \sum_{m,n} \Omega_{mn} \delta_{m'm} \delta_{nn'} = \Omega_{m'n'}, \quad (2.46)$$

as expected from the definition (2.42).

2.1 Projection Operators

Consider the familiar orthonormal basis $\{|i\rangle\}$ of V and choose one element $|m\rangle$ from the basis to form an operator P_m defined by

$$P_m \equiv |m\rangle \langle m|. \quad (2.47)$$

This operator maps any vector $|v\rangle \in V$ to a vector along $|, \rangle$. Indeed, acting on $|v\rangle$ it gives

$$P_m|v\rangle = |m\rangle \langle m|v\rangle \sim |m\rangle. \quad (2.48)$$

Comparing the above expression for P_m with (2.45) we see that in the chosen basis, P_n is represented by a matrix all of whose elements are zero, except for the (n, n) element $(P_n)_{nn}$ which is one:

$$P_n \longleftrightarrow \begin{pmatrix} 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \end{pmatrix}. \quad (2.49)$$

A hermitian operator P is said to be a *projection* operator if it satisfies the operator equation $PP = P$. This means that acting twice with a projection operator on a vector gives the same as acting once. The operator P_m is a projection operator since

$$P_m P_m = (|m\rangle\langle m|)(|m\rangle\langle m|) = |m\rangle\langle m|m\rangle\langle m| = |m\rangle\langle m|, \quad (2.50)$$

since $\langle m|m\rangle = 1$. The operator P_m is said to be a *rank one* projection operator since it projects to a one-dimensional subspace of V , the subspace generated by $|m\rangle$.

Using the basis vector $|m\rangle$ with $m \neq n$ we can define

$$P_{m,n} \equiv |m\rangle\langle m| + |n\rangle\langle n|. \quad (2.51)$$

Acting on any vector $|v\rangle \in V$, this operator gives us a vector in the subspace spanned by $|m\rangle$ and $|n\rangle$:

$$P_{m,n}|v\rangle = |m\rangle\langle m|v\rangle + |n\rangle\langle n|v\rangle. \quad (2.52)$$

Using the orthogonality of $|m\rangle$ and $|n\rangle$ we quickly find that $P_{m,n}P_{m,n} = P_{m,n}$ and therefore $P_{m,n}$ is a projector. It is a rank two projector, since it projects to a two-dimensional subspace of V , the subspace spanned by $|m\rangle$ and $|n\rangle$. Similarly, we can construct a rank three projector by adding an extra term $|k\rangle\langle k|$ with $k \neq m$ and $k \neq n$. If we include all basis vectors we would have the operator

$$P_{1,\dots,N} \equiv |1\rangle\langle 1| + |2\rangle\langle 2| + \dots + |N\rangle\langle N|. \quad (2.53)$$

As a matrix $P_{1,\dots,N}$ has a one on every element of the diagonal and a zero everywhere else. This is therefore the unit matrix, which represents the identity operator. Indeed we anticipated this in (1.29), and we thus write

$$\mathbf{1} = \sum_i |i\rangle\langle i|. \quad (2.54)$$

This is the completeness relation for the chosen orthonormal basis. This equation is sometimes called the ‘resolution’ of the identity.

Example. For the spin one-half system the unit operator can be written as a sum of two terms since the vector space is two dimensional. Using the orthonormal basis vectors $|+\rangle$ and $|-\rangle$ for spins along the positive and negative z directions, respectively, we have

$$\mathbf{1} = |+\rangle\langle +| + |-\rangle\langle -|. \quad (2.55)$$

Example. We can use the completeness relation to show that our formula (2.42) for matrix elements is consistent with matrix multiplication. Indeed for the product $\Omega_1\Omega_2$ of two operators we write

$$\begin{aligned} (\Omega_1\Omega_2)_{mn} &= \langle m|\Omega_1\Omega_2|n\rangle = \langle m|\Omega_1 \mathbf{1} \Omega_2|n\rangle \\ &= \langle m|\Omega_1 \left(\sum_{k=1}^N |k\rangle\langle k| \right) \Omega_2|n\rangle = \sum_{k=1}^N \langle m|\Omega_1|k\rangle\langle k|\Omega_2|n\rangle = \sum_{k=1}^N (\Omega_1)_{mk}(\Omega_2)_{kn}. \end{aligned} \quad (2.56)$$

This is the expected rule for the multiplication of the matrices corresponding to Ω_1 and Ω_2 .

2.2 Adjoint of a linear operator

A linear operator Ω on V is defined by its action on the vectors in V . We have noted that Ω can also be viewed as a linear operator on the dual space V^* . We defined the linear operator Ω^\dagger associated with Ω . In general Ω^\dagger by

$$\langle \Omega^\dagger u | v \rangle = \langle u | \Omega v \rangle \quad (2.57)$$

Flipping the order on the left-hand side we get

$$\langle v | \Omega^\dagger u \rangle^* = \langle u | \Omega v \rangle \quad (2.58)$$

Complex conjugating, and writing the operators more explicitly

$$\boxed{\langle v | \Omega^\dagger u \rangle = \langle u | \Omega | v \rangle^*, \quad \forall u, v.} \quad (2.59)$$

Flipping the two sides of (2.57) we also get

$$\langle v | \Omega^\dagger | u \rangle = \langle \Omega v | u \rangle \quad (2.60)$$

from which, taking the ket away, we learn that

$$\boxed{\langle v | \Omega^\dagger \equiv \langle \Omega v |.} \quad (2.61)$$

Another way to state the action of the operator Ω^\dagger is as follows. The linear operator Ω induces a map $|v\rangle \rightarrow |v'\rangle$ of vectors in V and, in fact, is defined by giving a complete list of these maps. The operator Ω^\dagger is defined as the one that induces the maps $\langle v | \rightarrow \langle v' |$ of the *corresponding* bras. Indeed,

$$\begin{aligned} |v'\rangle &= |\Omega v\rangle = \Omega |v\rangle, \\ \langle v'| &= \langle \Omega v | = \langle v | \Omega^\dagger \end{aligned} \quad (2.62)$$

The first line is just definitions. On the second line, the first equality is obtained by taking bras of the first equality on the first line. The second equality is just (2.61). We say it as

$$\boxed{\text{The bra associated with } \Omega |v\rangle \text{ is } \langle v | \Omega^\dagger.} \quad (2.63)$$

To see what hermiticity means at the level of matrix elements, we take u, v to be orthonormal basis vectors in (2.59)

$$\langle i|\Omega^\dagger|j\rangle = \langle j|\Omega|i\rangle^* \rightarrow (\Omega^\dagger)_{ij} = (\Omega_{ji})^*. \quad (2.64)$$

In matrix notation we have $\Omega^\dagger = (\Omega^t)^*$ where the superscript t denotes transposition.

Exercise. Show that $(\Omega_1\Omega_2)^\dagger = \Omega_2^\dagger\Omega_1^\dagger$ by taking matrix elements.

Exercise. Given an operator $\Omega = |a\rangle\langle b|$ for arbitrary vectors a, b , write a bra-ket expression for Ω^\dagger .

Solution: Acting with Ω on $|v\rangle$ and then taking the dual gives

$$\Omega|v\rangle = |a\rangle\langle b|v\rangle \rightarrow \langle v|\Omega^\dagger = \langle v|b\rangle\langle a|, \quad (2.65)$$

Since this equation is valid for any bra $\langle v|$ we read

$$\Omega^\dagger = |b\rangle\langle a|. \quad (2.66)$$

2.3 Hermitian and Unitary Operators

A linear operator Ω is said to be *hermitian* if it is equal to its adjoint:

$$\boxed{\text{Hermitian Operator: } \Omega^\dagger = \Omega.} \quad (2.67)$$

In quantum mechanics Hermitian operators are associated with observables. The eigenvalues of a Hermitian operator are the possible measured values of the observables. As we will show soon, the eigenvalues of a Hermitian operator are all real. An operator A is said to be *anti-hermitian* if $A^\dagger = -A$.

Exercise: Show that the commutator $[\Omega_1, \Omega_2]$ of two hermitian operators Ω_1 and Ω_2 is anti-hermitian.

There are a couple of equations that rewrite in useful ways the main property of Hermitian operators. Using $\Omega^\dagger = \Omega$ in (2.59) we find

$$\boxed{\text{If } \Omega \text{ is a Hermitian Operator: } \langle v|\Omega|u\rangle = \langle u|\Omega|v\rangle^*, \quad \forall u, v.} \quad (2.68)$$

It follows that the expectation value of a Hermitian operator in *any* state is real

$$\langle v|\Omega|v\rangle \text{ is real for any hermitian } \Omega. \quad (2.69)$$

Another neat form of the hermiticity condition is derived as follows:

$$\langle \Omega u|v\rangle = \langle u|\Omega^\dagger|v\rangle = \langle u|\Omega|v\rangle = \langle u|\Omega v\rangle, \quad (2.70)$$

so that all in all

$$\boxed{\text{Hermitian Operator: } \langle \Omega u|v\rangle = \langle u|\Omega v\rangle.} \quad (2.71)$$

In this expression we see that a hermitian operator moves freely from the bra to the ket (and viceversa).

Example: For wavefunction $f(x) \in \mathbb{C}$ we have written

$$\langle f|g \rangle = \int_{-\infty}^{\infty} (f(x))^* g(x) dx \quad (2.72)$$

For a Hermitian Ω we have $\langle \Omega f | g \rangle = \langle f | \Omega g \rangle$ or explicitly

$$\int_{-\infty}^{\infty} (\Omega f(x))^* g(x) dx = \int_{-\infty}^{\infty} (f(x))^* \Omega g(x) dx \quad (2.73)$$

Verify that the linear operator $\Omega = \frac{\hbar}{i} \frac{d}{dx}$ is hermitian when we restrict to functions that vanish at $\pm\infty$.

An operator U is said to be a unitary operator if U^\dagger is an inverse for U , that is, $U^\dagger U$ and UU^\dagger are both the identity operator:

$$U \text{ is a unitary operator: } U^\dagger U = UU^\dagger = \mathbf{1} \quad (2.74)$$

In finite dimensional vector spaces $U^\dagger U = \mathbf{1}$ implies $UU^\dagger = 1$, but this is not always the case for infinite dimensional vector spaces. A key property of unitary operators is that they preserve the norm of states. Indeed, assume that $|\psi'\rangle$ is obtained by the action of U on $|\psi\rangle$:

$$|\psi'\rangle = U|\psi\rangle \quad (2.75)$$

Taking the dual we have

$$\langle \psi' | = \langle \psi | U^\dagger, \quad (2.76)$$

and therefore

$$\langle \psi' | \psi' \rangle = \langle \psi | U^\dagger U | \psi \rangle = \langle \psi | \psi \rangle, \quad (2.77)$$

showing that $|\psi\rangle$ and $U|\psi\rangle$ are states with the same norm. More generally

$$\langle Ua | Ub \rangle = \langle a | U^\dagger U | b \rangle = \langle a | b \rangle. \quad (2.78)$$

Another important property of unitary operators is that acting on an orthonormal basis they give another orthonormal basis. To show this consider the orthonormal basis

$$|a_1\rangle, |a_2\rangle, \dots |a_N\rangle, \quad \langle a_i | a_j \rangle = \delta_{ij} \quad (2.79)$$

Acting with U we get

$$|Ua_1\rangle, |Ua_2\rangle, \dots |Ua_N\rangle, \quad (2.80)$$

To show that this is a basis we must prove that

$$\sum_i \beta_i |Ua_i\rangle = 0 \quad (2.81)$$

implies $\beta_i = 0$ for all i . Indeed, the above gives

$$\sum_i \beta_i |Ua_i\rangle = \sum_i \beta_i U|a_i\rangle = U \sum_i \beta_i |a_i\rangle = 0. \quad (2.82)$$

Acting with U^\dagger from the left we find that $\sum_i \beta_i |a_i\rangle = 0$ and, since the $|a_i\rangle$ form a basis, we get $\beta_i = 0$ for all i , as desired. The new basis is orthonormal because

$$\langle Ua_i|Ua_j\rangle = \langle a_i|U^\dagger U|a_j\rangle = \langle a_i|a_j\rangle = \delta_{ij}. \quad (2.83)$$

It follows from the above that the operator U can be written as

$$U = \sum_{i=1}^N |Ua_i\rangle\langle a_i|, \quad (2.84)$$

since

$$U|a_j\rangle = \sum_{i=1}^N |Ua_i\rangle\langle a_i|a_j\rangle = |Ua_j\rangle. \quad (2.85)$$

In fact for *any* unitary operator U in a vector space V there exist orthonormal bases $\{|a_i\rangle\}$ and $\{|b_i\rangle\}$ such that U can be written as

$$U = \sum_{i=1}^N |b_i\rangle\langle a_i|. \quad (2.86)$$

Indeed, this is just a rewriting of (2.84), with $|a_i\rangle$ any orthonormal basis and $|b_i\rangle = |Ua_i\rangle$.

Exercise: Verify that U in (2.86) satisfies $U^\dagger U = UU^\dagger = \mathbf{1}$.

Exercise: Prove that $\langle a_i|U|a_j\rangle = \langle b_i|U|b_j\rangle$.

3 Non-denumerable basis

In this section we describe the use of bras and kets for the position and momentum states of a particle moving on the real line $x \in \mathbb{R}$.

Let us begin with position. We will introduce position states $|x\rangle$ where the label x in the ket is the value of the position. Since x is a continuous variable and we position states $|x\rangle$ for all values of x to form a basis, we are dealing with an infinite basis that is not possible to label as $|1\rangle, |2\rangle, \dots$, it is a non-denumerable basis. So we have

$$\text{Basis states : } |x\rangle, \quad \forall x \in \mathbb{R}. \quad (3.87)$$

Basis states with different values of x are different vectors in the state space (a complex vector space, as always in quantum mechanics). Note here that the label on the ket is not a vector! So $|ax\rangle \neq a|x\rangle$, for any real $a \neq 1$. In particular $|-x\rangle \neq |x\rangle$ unless $x = 0$. For quantum mechanics in three dimensions, we have position states $|\vec{x}\rangle$. Here the label is a vector in a three-dimensional real vector space (our space!) while the ket is a vector in the infinite dimensional complex vector space of states of the theory.

Again something like $|\vec{x}_1 + \vec{x}_2\rangle$ has nothing to do with $|\vec{x}_1\rangle + |\vec{x}_2\rangle$. The $| \rangle$ enclosing the label of the position eigenstates plays a crucial role: it helps us see that object lives in an infinite dimensional complex vector space.

The inner product must be defined, so we will take

$$\langle x|y\rangle = \delta(x - y). \quad (3.88)$$

It follows that position states with different positions are orthogonal to each other. The norm of a position state is infinite: $\langle x|x\rangle = \delta(0) = \infty$, so these are not allowed states of particles. We visualize the state $|x\rangle$ as the state of a particle perfectly localized at x , but this is an idealization. We can easily construct normalizable states using superpositions of position states. We also have a completeness relation

$$\mathbf{1} = \int dx |x\rangle\langle x|. \quad (3.89)$$

This is consistent with our inner product above. Letting the above equation act on $|y\rangle$ we find an equality:

$$|y\rangle = \int dx |x\rangle\langle x|y\rangle = \int dx |x\rangle \delta(x - y) = |y\rangle. \quad (3.90)$$

The position operator \hat{x} is defined by its action on the position states. Not surprisingly we let

$$\hat{x}|x\rangle = x|x\rangle, \quad (3.91)$$

thus declaring that $|x\rangle$ are \hat{x} eigenstates with eigenvalue equal to the position x . We can also show that \hat{x} is a Hermitian operator by checking that \hat{x}^\dagger and \hat{x} have the same matrix elements:

$$\langle x_1|\hat{x}^\dagger|x_2\rangle = \langle x_2|\hat{x}|x_1\rangle^* = [x_1\delta(x_1 - x_2)]^* = x_2\delta(x_1 - x_2) = \langle x_1|\hat{x}|x_2\rangle. \quad (3.92)$$

We thus conclude that $\hat{x}^\dagger = \hat{x}$ and the bra associated with (3.91) is

$$\langle x|\hat{x} = x\langle x|. \quad (3.93)$$

Given the state $|\psi\rangle$ of a particle, we define the associated position-state wavefunction $\psi(x)$ by

$$\psi(x) \equiv \langle x|\psi\rangle \in \mathbb{C}. \quad (3.94)$$

This is sensible: $\langle x|\psi\rangle$ is a number that depends on the value of x , thus a function of x . We can now do a number of basic computations. First we write any state as a superposition of position eigenstates, by inserting $\mathbf{1}$ as in the completeness relation

$$|\psi\rangle = \mathbf{1}|\psi\rangle = \int dx |x\rangle\langle x|\psi\rangle = \int dx |x\rangle \psi(x). \quad (3.95)$$

As expected, $\psi(x)$ is the component of ψ along the state $|x\rangle$. Overlap of states can also be written in position space:

$$\langle\phi|\psi\rangle = \int dx \langle\phi|x\rangle\langle x|\psi\rangle = \int dx \phi^*(x)\psi(x). \quad (3.96)$$

Matrix elements involving \hat{x} are also easily evaluated

$$\langle \phi | \hat{x} | \psi \rangle = \langle \phi | \hat{x} \mathbf{1} | \psi \rangle = \int dx \langle \phi | \hat{x} | x \rangle \langle x | \psi \rangle = \int dx \langle \phi | x \rangle x \langle x | \psi \rangle = \int dx \phi^*(x) x \psi(x). \quad (3.97)$$

We now introduce momentum states $|p\rangle$ that are eigenstates of the momentum operator \hat{p} in complete analogy to the position states

$$\begin{aligned} \text{Basis states : } & |p\rangle, \quad \forall p \in \mathbb{R}. \\ \langle p' | p \rangle &= \delta(p - p'), \\ \mathbf{1} &= \int dp |p\rangle \langle p|, \\ \hat{p} |p\rangle &= p |p\rangle \end{aligned} \quad (3.98)$$

Just as for coordinate space we also have

$$\hat{p}^\dagger = \hat{p}, \quad \text{and} \quad \langle p | \hat{p} = p \langle p|. \quad (3.99)$$

In order to relate the two bases we need the value of the overlap $\langle x | p \rangle$. Since we interpret this as the wavefunction for a particle with momentum p we have from (6.39) of Chapter 1 that

$$\langle x | p \rangle = \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}. \quad (3.100)$$

The normalization was adjusted properly to be compatible with the completeness relations. Indeed, for example, consider the $\langle p' | p \rangle$ overlap and use the completeness in x to evaluate it

$$\langle p' | p \rangle = \int dx \langle p' | x \rangle \langle x | p \rangle = \frac{1}{2\pi\hbar} \int dx e^{i(p-p')x/\hbar} = \frac{1}{2\pi} \int du e^{i(p-p')u}, \quad (3.101)$$

where we let $u = x/\hbar$ in the last step. We claim that the last integral is precisely the integral representation of the delta function $\delta(p - p')$:

$$\frac{1}{2\pi} \int du e^{i(p-p')u} = \delta(p - p'). \quad (3.102)$$

This, then gives the correct value for the overlap $\langle p | p' \rangle$, as we claimed. The integral (3.102) can be justified using the fact that the functions

$$f_n(x) \equiv \frac{1}{\sqrt{L}} \exp\left(\frac{2\pi i n x}{L}\right), \quad (3.103)$$

form a complete orthonormal set of functions over the interval $x \in [-L/2, L/2]$. Completeness then means that

$$\sum_{n \in \mathbb{Z}} f_n^*(x) f_n(x') = \delta(x - x'). \quad (3.104)$$

We thus have

$$\sum_{n \in \mathbb{Z}} \frac{1}{L} \exp\left(2\pi i \frac{n}{L}(x - x')\right) = \delta(x - x'). \quad (3.105)$$

In the limit as L goes to infinity the above sum can be written as an integral since the exponential is a very slowly varying function of $n \in \mathbb{Z}$. Since $\Delta n = 1$ with $u = 2\pi n/L$ we have $\Delta u = 2\pi/L \ll 1$ and then

$$\sum_{n \in \mathbb{Z}} \frac{1}{L} \exp\left(2\pi i \frac{n}{L}(x - x')\right) = \sum_u \frac{\Delta u}{2\pi} \exp\left(i u(x - x')\right) \rightarrow \frac{1}{2\pi} \int du e^{iu(x-x')} , \quad (3.106)$$

and back in (3.105) we have justified (3.102).

We can now ask: What is $\langle p|\psi\rangle$? We compute

$$\langle p|\psi\rangle = \int dx \langle p|x\rangle \langle x|\psi\rangle = \frac{1}{\sqrt{2\pi\hbar}} \int dx e^{-ipx/\hbar} \psi(x) = \tilde{\psi}(p) , \quad (3.107)$$

which is the Fourier transform of $\psi(x)$, as defined in (6.41) of Chapter 1. Thus the Fourier transform of $\psi(x)$ is the wavefunction in the momentum representation.

It is useful to know how to evaluate $\langle x|\hat{p}|\psi\rangle$. We do it by inserting a complete set of momentum states:

$$\langle x|\hat{p}|\psi\rangle = \int dp \langle x|p\rangle \langle p|\hat{p}|\psi\rangle = \int dp (p\langle x|p\rangle) \langle p|\psi\rangle \quad (3.108)$$

Now we notice that

$$p\langle x|p\rangle = \frac{\hbar}{i} \frac{d}{dx} \langle x|p\rangle \quad (3.109)$$

and thus

$$\langle x|\hat{p}|\psi\rangle = \int dp \left(\frac{\hbar}{i} \frac{d}{dx} \langle x|p\rangle \right) \langle p|\psi\rangle . \quad (3.110)$$

The derivative can be moved out of the integral, since no other part of the integrand depends on x :

$$\langle x|\hat{p}|\psi\rangle = \frac{\hbar}{i} \frac{d}{dx} \int dp \langle x|p\rangle \langle p|\psi\rangle \quad (3.111)$$

The completeness sum is now trivial and can be discarded to obtain

$$\langle x|\hat{p}|\psi\rangle = \frac{\hbar}{i} \frac{d}{dx} \langle x|\psi\rangle = \frac{\hbar}{i} \frac{d}{dx} \psi(x) . \quad (3.112)$$

Exercise. Show that

$$\langle p|\hat{x}|\psi\rangle = i\hbar \frac{d}{dp} \tilde{\psi}(p) . \quad (3.113)$$

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

WAVE MECHANICS

B. Zwiebach
September 13, 2013

Contents

1 The Schrödinger equation	1
2 Stationary Solutions	4
3 Properties of energy eigenstates in one dimension	10
4 The nature of the spectrum	12
5 Variational Principle	18
6 Position and momentum	22

1 The Schrödinger equation

In classical mechanics the motion of a particle is usually described using the time-dependent position $\vec{x}(t)$ as the dynamical variable. In wave mechanics the dynamical variable is a wavefunction. This wavefunction depends on position and on time and it is a complex number – it belongs to the complex numbers \mathbb{C} (we denote the real numbers by \mathbb{R}). When all three dimensions of space are relevant we write the wavefunction as

$$\Psi(\vec{x}, t) \in \mathbb{C}. \quad (1.1)$$

When only one spatial dimension is relevant we write it as $\Psi(x, t) \in \mathbb{C}$. The wavefunction satisfies the Schrödinger equation. For one-dimensional space we write

$$i\hbar \frac{\partial \Psi}{\partial t}(x, t) = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x, t) \right) \Psi(x, t). \quad (1.2)$$

This is the equation for a (non-relativistic) particle of mass m moving along the x axis while acted by the potential $V(x, t) \in \mathbb{R}$. It is clear from this equation that the wavefunction must be complex: if it were real, the right-hand side of (1.2) would be real while the left-hand side would be imaginary, due to the explicit factor of i .

Let us make two important remarks:

1. The Schrödinger equation is a *first order* differential equation in time. This means that if we prescribe the wavefunction $\Psi(x, t_0)$ for all of space at an arbitrary initial time t_0 , the wavefunction is determined for all times.
2. The Schrödinger equation is a *linear* equation for Ψ : if Ψ_1 and Ψ_2 are solutions so is $a_1\Psi_1 + a_2\Psi_2$ with a_1 and a_2 arbitrary *complex numbers*.

Given a complex number $z = a + ib$, $a, b \in \mathbb{R}$, its complex conjugate is $z^* = a - ib$. Let $|z|$ denote the norm or length of the complex number z . The norm is a positive number (thus real!) and it is given by $|z| = \sqrt{a^2 + b^2}$. If the norm of a complex number is zero, the complex number is zero. You can quickly verify that

$$|z|^2 = zz^*. \quad (1.3)$$

For a wavefunction $\Psi(x, t)$ its complex conjugate $(\Psi(x, t))^*$ will be usually written as $\Psi^*(x, t)$.

We define the probability density $P(x, t)$, also denoted as $\rho(x, t)$, as the norm-squared of the wavefunction:

$$P(x, t) = \rho(x, t) \equiv \Psi^*(x, t)\Psi(x, t) = |\Psi(x, t)|^2. \quad (1.4)$$

This probability density so defined is positive. The physical interpretation of the wavefunction arises because we declare that

$P(x, t) dx$ is the probability to find the particle in the interval $[x, x + dx]$ at time t .

(1.5)

This interpretation requires a *normalized* wavefunction, namely, the wavefunction used above must satisfy, for all times,

$$\int_{-\infty}^{\infty} dx |\Psi(x, t)|^2 = 1, \quad \forall t. \quad (1.6)$$

By integrating over space, the left-hand adds up the probabilities that the particle be found in all of the tiny intervals dx that comprise the real line. Since the particle must be found somewhere this sum must be equal to one.

Suppose you are handed a wavefunction that is normalized at time t_0 :

$$\int_{-\infty}^{\infty} dx |\Psi(x, t_0)|^2 = 1, \quad \forall t. \quad (1.7)$$

As mentioned above, knowledge of the wavefunction at one time implies, via the Schrödinger equation, knowledge for all times. The Schrödinger equation must guarantee that the wavefunction remains normalized for all times. Proving this is a good exercise:

Exercise 1. Show that the Schrödinger equation implies that the norm of the wavefunction does not change in time:

$$\frac{d}{dt} \int_{-\infty}^{\infty} dx |\Psi(x, t)|^2 = 0. \quad (1.8)$$

You will have to use both the Schrödinger equation and its complex-conjugate version. Moreover you will have to use $\Psi(x, t) \rightarrow 0$ as $|x| \rightarrow \infty$, which is true, as no normalizable wavefunction can take a non-zero value as $|x| \rightarrow \infty$. While generally the derivative $\frac{\partial}{\partial x} \Psi$ also goes to zero as $|x| \rightarrow \infty$ you only need to assume that it remains bounded.

- - - - -

Associated to the probability density $\rho(x, t) = \Psi^* \Psi$ there is a **probability current** $J(x, t)$ that characterizes the flow of probability and is given by

$$J(x, t) = \frac{\hbar}{m} \text{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right). \quad (1.9)$$

The analogy in electromagnetism is useful. There we have the current density vector \vec{J} and the charge density ρ . The statement of charge conservation is the differential relation

$$\nabla \cdot \vec{J} + \frac{\partial \rho}{\partial t} = 0. \quad (1.10)$$

This equation applied to a fixed volume V implies that the rate of change of the enclosed charge $Q_V(t)$ is only due to the flux of \vec{J} across the surface S that bounds the volume:

$$\frac{dQ_V}{dt}(t) = - \oint_S \vec{J} \cdot d\vec{a}. \quad (1.11)$$

Make sure you know how to get this equation from (1.10)! While the probability current in more than one spatial dimension is also a vector, in our present one-dimensional case, it has just one component. The conservation equation is the analog of (1.10):

$$\frac{\partial J}{\partial x} + \frac{\partial \rho}{\partial t} = 0. \quad (1.12)$$

You can check that this equation holds using the above formula for $J(x, t)$, the formula for $\rho(x, t)$, and the Schrödinger equation. The integral version is formulated by first defining the probability $P_{ab}(t)$ of finding the particle in the interval $x \in [a, b]$

$$P_{ab}(t) \equiv \int_a^b dx |\Psi(x, t)|^2 = \int_a^b dx \rho(x, t). \quad (1.13)$$

You can then quickly show that

$$\frac{dP_{ab}}{dt}(t) = J(a, t) - J(b, t). \quad (1.14)$$

Here $J(a, t)$ denotes the rate at which probability flows in (in units of one over time) at the left boundary of the interval, while $J(b, t)$ denotes the rate at which probability flows out at the right boundary of the interval.

It is sometimes easier to work with wavefunctions that are not normalized. The normalization can be performed if needed. We will thus refer to wavefunctions in general without assuming normalization, otherwise we will call them *normalized* wavefunction. In this spirit, two wavefunctions Ψ_1 and Ψ_2 solving the Schrödinger equation are *declared* to be physically equivalent if they differ by multiplication by a complex *number*. Using the symbol \sim for equivalence, we write

$$\Psi_1 \sim \Psi_2 \iff \Psi_1(x, t) = \alpha \Psi_2(x, t), \quad \alpha \in \mathbb{C}. \quad (1.15)$$

If the wavefunctions Ψ_1 and Ψ_2 are normalized they are equivalent if they differ by an overall constant phase:

$$\text{Normalized wavefunctions: } \Psi_1 \sim \Psi_2 \iff \Psi_1(x, t) = e^{i\theta} \Psi_2(x, t), \quad \theta \in \mathbb{R}. \quad (1.16)$$

2 Stationary Solutions

In a large class of problems the Schrödinger potential $V(x, t)$ has no time dependence and it is simply a function $V(x)$ of position. We focus on that case now. The Schrödinger equation (1.2) can be written more briefly as

$$i\hbar \frac{\partial \Psi}{\partial t}(x, t) = \hat{H} \Psi(x, t), \quad (2.1)$$

where we have introduced the Hamiltonian operator \hat{H} :

$$\hat{H} \equiv -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x). \quad (2.2)$$

\hat{H} is an operator in the sense that it acts on functions of x and t to give functions of x and t : it acts on the space of complex functions, a space that contains wavefunctions. Note that $V(x)$ acts just by multiplication. Note that the operator \hat{H} is time independent – it does not involve time at all.

A **stationary** state of energy $E \in \mathbb{R}$ is a state $\Psi(x, t)$ that takes the form

$$\Psi(x, t) = e^{-iEt/\hbar} \psi(x), \quad (2.3)$$

where $\psi(x) \in \mathbb{C}$ is a function of x only that solves an equation that will be discussed below. All the time dependence of the stationary state is carried by the exponential prefactor. Such a state is called stationary because physical observables of the state are actually time independent. Consider, for example, the norm of the state. We see that the time dependence drops out

$$P(x, t) = \Psi^*(x, t) \Psi(x, t) = e^{+iEt/\hbar} \psi^*(x) e^{-iEt/\hbar} \psi(x) = \psi^*(x) \psi(x) = |\psi(x)|^2. \quad (2.4)$$

Had the energy E been a complex number $E = E_0 - i\Gamma$, with E_0 and Γ real, the time dependence would not drop out:

$$\begin{aligned} P(x, t) &= \Psi^*(x, t) \Psi(x, t) = e^{+iE^*t/\hbar} \psi^*(x) e^{-iEt/\hbar} \psi(x) \\ &= e^{i(E^*-E)t/\hbar} \psi^*(x) \psi(x) = e^{-2\Gamma t/\hbar} |\psi(x)|^2. \end{aligned} \quad (2.5)$$

This kind of state is not acceptable: the normalization cannot be preserved in time.

Let us consider the equation that $\psi(x)$ must satisfy. Plugging (2.3) into (2.1) we find

$$i\hbar \frac{\partial}{\partial t} e^{-iEt/\hbar} \psi(x) = \hat{H} e^{-iEt/\hbar} \psi(x), \quad (2.6)$$

The time derivative on the left-hand side only acts on the exponential and the \hat{H} operator on the right-hand side can be moved through the exponential (it commutes with it!) so we get

$$i\hbar \left(\frac{\partial}{\partial t} e^{-iEt/\hbar} \right) \psi(x) = e^{-iEt/\hbar} \hat{H} \psi(x). \quad (2.7)$$

Taking the derivative and canceling the exponentials we get

$$i\hbar \left(-i \frac{E}{\hbar} \right) \psi(x) = \hat{H} \psi(x), \quad (2.8)$$

which we write as

$$\boxed{\hat{H} \psi(x) = E \psi(x)}. \quad (2.9)$$

Recalling the expression for the Hamiltonian we have

$$\boxed{\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right) \psi(x) = E \psi(x)}. \quad (2.10)$$

Note that the derivatives along x need not be denoted as partial derivatives since the functions they act on have no other argument except x . Using primes to denote derivatives with respect to the argument, the above equation is

$$\boxed{-\frac{\hbar^2}{2m} \psi''(x) + V(x) \psi(x) = E \psi(x)}. \quad (2.11)$$

This is the equation for ψ that makes $\Psi(x, t) = e^{-iEt/\hbar} \psi(x)$ a stationary state of energy E . Any of the three boxed equations above is referred to as the **time-independent** Schrödinger equation.

Since the time-independent Schrödinger equation is a second-order differential equation in x , a solution is completely determined by the value of ψ and ψ' at some point x_0 ($\neq \infty$). If $V(x)$

is nowhere infinite, $\psi = \psi' = 0$ at some point implies $\psi = 0$ everywhere. Alternatively, if we know the solution for *any* size x -interval, the full solution is fully determined. A full solution means finding all the values E for which acceptable solutions $\psi(x)$ exist and, of course, finding those solutions for each E .

A solution $\psi(x)$ associated with an energy E is called an **energy eigenstate** of energy E . The set of all allowed values of E is called the **spectrum** of the Hamiltonian \hat{H} . A **degeneracy** in the spectrum occurs when there is more than one solution $\psi(x)$ for a given value of the energy.

The solutions depend on the properties of the potential $V(x)$. We will consider potentials $V(x)$ that can fail to be continuous (but are piece-wise continuous, like the finite square well) and can fail to be bounded (like the potential for the harmonic oscillator). We allow delta function contributions in the potential but do not allow worse singularities, such as squares or derivatives of delta functions. We allow hard walls, as in the infinite square-well.

On the wavefunction we impose the following regularity condition:

$$\psi(x) \text{ is continuous and bounded and its derivative } \psi'(x) \text{ is bounded.} \quad (2.12)$$

We do not impose the requirement that $\psi(x)$ be normalizable. This would be too restrictive. There are energy eigenstates that are not normalizable. Momentum eigenstates of a free particle are also not normalizable. Solutions for which ψ is not normalizable do not have a direct physical interpretation, but are very useful: suitable superpositions of them give normalizable solutions that can represent a particle.

In the spectrum of a Hamiltonian, localized energy eigenstates are particularly important. This motivates the definition:

$$\text{An energy eigenstate } \psi(x) \text{ is a } \mathbf{bound} \text{ state if } \psi(x) \rightarrow 0 \text{ when } |x| \rightarrow \infty. \quad (2.13)$$

Since a normalizable eigenstate must have a wavefunction that vanishes as $|x| \rightarrow \infty$, a bound state is just a normalizable eigenstate.

The eigenstates of \hat{H} provide a useful set of functions. Let us denote the possible energies by E_n with $n = 1, 2, \dots$, ordered as follows

$$E_1 \leq E_2 \leq E_3 \leq \dots \quad (2.14)$$

and let the corresponding eigenstates be $\psi_n(x)$, with

$$\hat{H}\psi_n(x) = E_n \psi_n(x), \quad n \geq 1. \quad (2.15)$$

For simplicity we discuss the case when the spectrum is denumerable so that, as above, we can label the states and energies with the integer label n . In general a potential $V(x)$ can result in a

spectrum that contains a discrete part and a continuous part. The discrete part is denumerable but the continuous part is not. The formulae we will write below require some modification when there spectrum contains a continuous part. The eigenstates of the continuum spectrum are not normalizable.

It is a known result about differential equations that for rather general potentials the \hat{H} eigenstates $\psi_n(x)$ can be chosen to be orthonormal. What does it mean for two functions to be orthogonal? Orthogonal vectors have a vanishing dot product, where the dot product is a (clever) rule to obtain a single number from two vectors. For two functions f_1 and f_2 an *inner* product can be defined by integrating the product function $f_1 f_2$ over all x , thus giving us a number. Since our functions are complex valued, a small modification is needed: the inner product of f_1 and f_2 is taken to be $\int f_1^* f_2$. The functions f_1 and f_2 are orthogonal if this integral vanishes. An orthonormal set of functions is one in which each function is orthogonal to all others, while its inner product with itself gives one (this requires the complex conjugation in the definition, can you see that?). As a result, orthonormality means that

$$\boxed{\text{Orthonormality: } \int_{-\infty}^{\infty} dx \psi_m^*(x) \psi_n(x) = \delta_{m,n}.} \quad (2.16)$$

Recall that the Kronecker delta $\delta_{m,n}$ is defined to be zero if $m \neq n$ and one otherwise.

The energy eigenstates are also *complete* in the sense that any reasonable (see (2.12) wavefunction $\psi(x)$ can be expanded as a superposition of energy eigenstates. Namely, there exist complex numbers b_n such that

$$\psi(x) = \sum_{n=1}^{\infty} b_n \psi_n(x), \quad b_n \in \mathbb{C}. \quad (2.17)$$

This is a very powerful statement: it means that if the energy eigenstates are known, the general solution of the Schrödinger equation is known. Indeed assume that the wavefunction at time equal zero is the $\psi(x)$ above. Then we have

$$\boxed{\Psi(x, t=0) = \psi(x) = \sum_{n=1}^{\infty} b_n \psi_n(x).} \quad (2.18)$$

If this wavefunction is normalized then we have

$$\int_{-\infty}^{\infty} dx \psi^*(x) \psi(x) = 1 \rightarrow \sum_{n=1}^{\infty} |b_n|^2 = 1. \quad (2.19)$$

We now claim that the wavefunction at all times can be written down immediately

$$\boxed{\Psi(x, t) = \sum_{n=1}^{\infty} b_n e^{-iE_n t/\hbar} \psi_n(x).} \quad (2.20)$$

To prove that this is the solution we first note that we have produced a solution to the Schrödinger equation: this follows by linearity because each term in the above sum is a solution (a stationary state). Second, the solution reduces for $t = 0$ to the correct value $\Psi(x, t = 0)$ in (2.18). By the first remark below (1.2) this is all that is needed.

It should be emphasized that the superposition of stationary states is generally *not* a stationary state. The expansion coefficients b_n used above can be calculated explicitly if we know the energy eigenstates. Indeed using (2.16) and (2.17) a one-line computation (do it!) gives

$$b_n = \int_{-\infty}^{\infty} dx \psi_n^*(x) \psi(x). \quad (2.21)$$

A curious identity can be derived by substituting this result back into (2.17):

$$\psi(x) = \sum_{n=1}^{\infty} \left(\int_{-\infty}^{\infty} dx' \psi_n^*(x') \psi(x') \right) \psi_n(x) = \int_{-\infty}^{\infty} dx' \left(\sum_{n=1}^{\infty} \psi_n^*(x') \psi_n(x) \right) \psi(x'), \quad (2.22)$$

where we interchanged the order of integration and summation (a safe operation in most cases!). The above equation is of the form

$$f(x) = \int_{-\infty}^{\infty} dx' K(x', x) f(x') \quad (2.23)$$

and is supposed to hold for any function $f(x)$. It is intuitively clear that that $K(x', x)$ must vanish for $x' \neq x$ for otherwise we could cook up a contradiction by choosing a peculiar function $f(x)$. Taking $f(x) = \delta(x - x_0)$ the equation gives

$$\delta(x - x_0) = \int dx' K(x', x) \delta(x' - x_0) = K(x_0, x). \quad (2.24)$$

We therefore conclude that $K(x', x) = \delta(x - x')$ (recall that $\delta(x) = \delta(-x)$). Back in (2.22) we thus find

$$\text{Completeness: } \sum_{n=1}^{\infty} \psi_n^*(x') \psi_n(x) = \delta(x - x'). \quad (2.25)$$

Let us compare the completeness relation above with the orthonormality relation (2.16). In the completeness relation we set equal the labels of the eigenfunctions and sum over them while keeping the two position arguments fixed. In the orthogonality relation we set equal the position arguments of the eigenfunctions and integrate (sum) over them while keeping the two labels fixed. On the right-hand sides we find “delta functions”: a Kronecker delta setting equal the two labels in the orthonormality relation and a true delta function setting equal the two positions in the completeness relation. The two relations are obtained from each other by exchange of labels: position labels and energy labels. This is a neat duality!

It is fun to calculate the expectation value of the Hamiltonian in the solution $\Psi(x, t)$ in (2.20). For arbitrary time-independent operators \hat{A} one defines the (generally) time-dependent expectation value on a *normalized* state Ψ by

$$\langle \hat{A} \rangle_{\Psi}(t) \equiv \int_{-\infty}^{\infty} dx \Psi^*(x, t)(\hat{A}\Psi(x, t)) . \quad (2.26)$$

What happens when we take the operator to be \hat{H} ? Using (2.20) twice, we get

$$\begin{aligned} \langle \hat{H} \rangle_{\Psi}(t) &= \int_{-\infty}^{\infty} dx \Psi^*(x, t)(\hat{H}\Psi(x, t)) \\ &= \sum_{n,n'} \int_{-\infty}^{\infty} dx b_n^* e^{iE_n t/\hbar} \psi_n^*(x) b_{n'} e^{-iE'_{n'} t/\hbar} \hat{H} \psi_{n'}(x) \\ &= \sum_{n,n'} b_n^* b_{n'} E_{n'} e^{i(E_n - E_{n'})t/\hbar} \int_{-\infty}^{\infty} dx \psi_n^*(x) \psi_{n'}(x) \\ &= \sum_{n,n'} b_n^* b_{n'} E_{n'} e^{i(E_n - E_{n'})t/\hbar} \delta_{n,n'} , \end{aligned} \quad (2.27)$$

so that we get

$$\langle \hat{H} \rangle_{\Psi}(t) = \sum_{n=1}^{\infty} |b_n|^2 E_n . \quad (2.28)$$

The expectation value of the Hamiltonian is time-independent: this is the quantum version of energy conservation. This is the expected value of the energy: a weighted sum of the possible energies with weights the norm-squared of the expansion coefficients.

If the wavefunction $\Psi(x, t)$ is not normalized but is normalizable, then the wavefunction

$$\frac{\Psi(x, t)}{\sqrt{\int dx \Psi^* \Psi}} \quad (2.29)$$

is normalized. We can thus use this normalized wavefunction in the definition on $\langle \hat{A} \rangle$ to find the expectation value is given by

$$\langle \hat{A} \rangle_{\Psi}(t) \equiv \frac{\int_{-\infty}^{\infty} dx \Psi^*(x, t)(\hat{A}\Psi(x, t))}{\int dx \Psi^*(x, t)\Psi(x, t)} . \quad (2.30)$$

This formula can be used for any normalizable Ψ . If the Ψ is normalized the formula reduces to the earlier expression for $\langle \hat{A} \rangle$.

Another operator often used to explore the physics of states is the momentum operator \hat{p} . Acting on wavefunctions that depend on a coordinate x it takes the form of a differential operator:

$$\hat{p} \equiv \frac{\hbar}{i} \frac{\partial}{\partial x} . \quad (2.31)$$

3 Properties of energy eigenstates in one dimension

In order to simplify our notation we rewrite the time-independent Schrödinger equation (2.10) as follows

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - V(x))\psi = 0. \quad (3.1)$$

We then define energy-like quantities \mathcal{E} and \mathcal{V} using a common rescaling factor:

$$\mathcal{E} \equiv \frac{2m}{\hbar^2}E, \quad \mathcal{V}(x) \equiv \frac{2m}{\hbar^2}V(x). \quad (3.2)$$

With this the Schrödinger equation (3.1) becomes

$$\psi'' + (\mathcal{E} - \mathcal{V}(x))\psi = 0. \quad (3.3)$$

We are now ready to consider a basic result: in a one-dimensional potential there cannot be two or more bound states for any given energy.

Theorem 1. There is no degeneracy for bound states in one-dimensional potentials.

Proof. Suppose there is such degeneracy so that there are $\psi_1(x)$ and $\psi_2(x)$, different from each other and both corresponding to the same energy E , thus same value of \mathcal{E} . If so, we have that the following equations hold

$$\begin{aligned} \psi_1'' + (\mathcal{E} - \mathcal{V}(x))\psi_1 &= 0, \\ \psi_2'' + (\mathcal{E} - \mathcal{V}(x))\psi_2 &= 0. \end{aligned} \quad (3.4)$$

Multiplying the top equation by ψ_2 and the bottom one by ψ_1 and subtracting them we find

$$\psi_2\psi_1'' - \psi_1\psi_2'' = 0. \quad (3.5)$$

The left-hand side is actually a derivative

$$(\psi_2\psi_1' - \psi_1\psi_2')' = 0. \quad (3.6)$$

It follows from this that the expression inside the parenthesis must be a constant c ,

$$\psi_2\psi_1' - \psi_1\psi_2' = c. \quad (3.7)$$

The constant can be evaluated by examining the left-hand side for $|x| \rightarrow \infty$. We then have that $\psi_1 \rightarrow 0$ and $\psi_2 \rightarrow 0$, since they are bound states, while the derivatives are bounded, as assumed in (2.12). It follows that the left-hand side vanishes as $|x| \rightarrow \infty$ and therefore $c = 0$. We thus have

$$\psi_2\psi_1' = \psi_1\psi_2' \quad \rightarrow \quad \frac{\psi_1'}{\psi_1} = \frac{\psi_2'}{\psi_2} \quad \rightarrow \quad \frac{d}{dx}(\ln \psi_1 - \ln \psi_2) = 0. \quad (3.8)$$

This implies that we have for some constant c'

$$\ln \psi_1 = \ln \psi_2 + \ln c' \quad \rightarrow \quad \psi_1(x) = c' \psi_2(x). \quad (3.9)$$

We have thus shown that the wavefunctions ψ_1 and ψ_2 are equivalent. In contradiction with the initial assumption, they are the same energy eigenstate. This concludes the proof.

For our second theorem we show that the reality of \mathcal{V} allows us to work with real wavefunctions.

Theorem 2. The energy eigenstates $\psi(x)$ can be chosen to be real.

Proof. Consider our main equation and a possibly complex wavefunction that correspond

$$\psi'' + (\mathcal{E} - \mathcal{V}(x))\psi = 0, \quad (3.10)$$

Since $(\psi'')^* = (\psi^*)''$ the complex conjugation of the above equation gives

$$(\psi^*)'' + (\mathcal{E} - \mathcal{V}(x))\psi^* = 0. \quad (3.11)$$

So ψ^* if different from ψ defines a degenerate solution. By superposition we can then get two real (degenerate) solutions

$$\psi_r \equiv \frac{1}{2}(\psi + \psi^*), \quad \psi_{im} \equiv \frac{1}{2i}(\psi - \psi^*). \quad (3.12)$$

These are, of course, the real and imaginary parts of ψ .

If we are dealing with bound states of one-dimensional potentials more can be said: any such solution is, up to a phase, *equal* to a real solution. Indeed, the absence of degenerate bound states means that the two real solutions ψ_r and ψ_{im} must be equal up to a constant that can only be real:

$$\psi_{im} = c \psi_r, \quad \text{with } c \in \mathbb{R} \quad (3.13)$$

It then follows that $\psi = \psi_r + i\psi_{im} = (1 + ic)\psi_r$. Writing $1 + ic = \sqrt{1 + c^2} e^{i\beta}$ with real beta, shows that ψ is, up to a phase β , equal to a real solution.

Theorem 3. If the potential is an even function of x : $V(-x) = V(x)$ the eigenstates can be chosen to be even or odd under $x \rightarrow -x$.

Proof. Again, we begin with our main equation

$$\psi''(x) + (\mathcal{E} - \mathcal{V}(x))\psi(x) = 0. \quad (3.14)$$

Recall that primes denote here derivative with respect to the argument, so $\psi''(x)$ means the function “second-derivative-of- ψ ” evaluated at x . Similarly $\psi''(-x)$ means the function “second-derivative-of- ψ ” evaluated at $-x$. Thus we can change x for $-x$ with impunity in the above equation getting

$$\psi''(-x) + (\mathcal{E} - \mathcal{V}(x))\psi(-x) = 0, \quad (3.15)$$

where we used that V , and thus \mathcal{V} , is even. We now want to make clear that the above equation implies that $\psi(-x)$ is another solution of the Schrödinger equation with the same energy. For this let us define

$$\varphi(x) \equiv \psi(-x) \rightarrow \frac{d}{dx}\varphi(x) = \psi'(-x) \cdot (-1). \quad (3.16)$$

Taking a second derivative and using (3.15)

$$\frac{d^2}{dx^2}\varphi(x) = \psi''(-x) = -(\mathcal{E} - \mathcal{V}(x))\varphi(x), \quad (3.17)$$

so that indeed $\varphi(x) = \psi(-x)$ provides a degenerate solution to the Schrödinger equation:

$$\frac{d^2}{dx^2}\varphi(x) + (\mathcal{E} - \mathcal{V}(x))\varphi(x) = 0. \quad (3.18)$$

Equipped with the degenerate solutions $\psi(x)$ and $\psi(-x)$ we can now form symmetric (s) and antisymmetric (a) combinations that are, respectively, even and odd under $x \rightarrow -x$:

$$\psi_s(x) \equiv \frac{1}{2}(\psi(x) + \psi(-x)), \quad \psi_a(x) \equiv \frac{1}{2}(\psi(x) - \psi(-x)). \quad (3.19)$$

These are the solutions claimed to exist in Theorem 3.

Again, if we focus on bound states of one-dimensional potentials the absence of degeneracy implies that $\psi(x)$ and $\psi(-x)$ must be the same solution. Because of Theorem 2 we can choose $\psi(x)$ to be real and thus we must have

$$\psi(-x) = c\psi(x), \quad \text{with } c \in \mathbb{R}. \quad (3.20)$$

Letting $x \rightarrow -x$ in the above equation we get $\psi(x) = c\psi(-x) = c^2\psi(x)$ from which we learn that $c^2 = 1$. The only possibilities are $c = \pm 1$. So $\psi(x)$ is *automatically* even or odd under $x \rightarrow -x$. Any one-dimensional bound state solution with an even potential *must be* either even or odd under $x \rightarrow -x$.

4 The nature of the spectrum

Consider the time-independent Schrödinger equation written as

$$\psi'' = -\frac{2m}{\hbar^2}(E - V(x))\psi. \quad (4.1)$$

We always have that $\psi(x)$ is continuous, otherwise ψ'' has singularities worse than delta functions and we would require potentials $V(x)$ that are worse than delta functions – something we will not consider. Consider now three possibilities concerning the potential:

1. $V(x)$ is continuous. In this case the continuity of $\psi(x)$ and (4.1) imply ψ'' is also continuous. This requires ψ' continuous.

2. $V(x)$ has finite jumps. In this case ψ'' has finite jumps (it is the multiplication of ψ with no jumps times V with jumps). But then ψ' can have no jumps (it is continuous, with non-continuous derivative).
3. $V(x)$ contains delta functions. In this case ψ'' also contains delta functions (it is the multiplication of the continuous ψ times a delta function in V). Thus ψ' has finite jumps.
4. $V(x)$ contains a hard wall. A potential that is finite immediately to the left of $x = a$ and becomes infinite for $x > a$ is said to have a hard wall at $x = a$. In such a case, the wavefunction will vanish for $x \geq a$. The slope ψ' will be finite as $x \rightarrow a$ from the left, and will vanish for $x > a$. Thus ψ' is discontinuous at the wall.

In conclusion

Both ψ and ψ' are continuous unless the potential has delta functions or hard walls in which cases ψ' may have finite jumps. (4.2)

The origin of the discrete and continuous spectrum can be seen from simple examples. We have three situations to discuss, as shown in Figure 1 as (a), (b), and (c). We will consider the number of parameters needed to write a solution and the number of constraints due to boundary conditions. Without loss of generality we can consider real solutions, and therefore the parameters will be real.

- (a) Here the energy E is below the potential far to the left and far to the right, but not in the middle. On the left the solution must be a decaying exponential $\alpha_1 \exp(-\kappa|x|)$, where α_1 is a constant to be determined and κ is known if the energy E is known. So thus far we got one unknown constant α_1 . In the middle region where $E > V$ the solution is oscillatory $\alpha_2 \cos kx + \alpha_3 \sin kx$, with two unknown constants α_2 and α_3 , and k determined if E is known. Finally to the right we have a solution $\alpha_4 \exp(-\kappa x)$ since the wavefunction must vanish as $x \rightarrow \infty$. So we got four (real) unknown constants α_i , $i = 1, 2, 3, 4$. Since ψ and $c\psi$ are the same solution we can scale the solution and thus we only have three unknown constants to determine. There are, however, four constraints from boundary conditions: the continuity of ψ and ψ' at each of the two interfaces. With three coefficients and four conditions we cannot expect a solution to exist. If we view the energy E , however, as unknown, then we have four unknowns and four conditions. Indeed solutions exist for discrete values of the energy. We get a discrete spectrum.
- (b) Here we have one unknown constant for the solution to the left of the interface (multiplying a decaying exponential) and two unknown constants for the oscillatory solution to the right of the interface, for a total of three unknowns, or just two unknowns once the overall scale

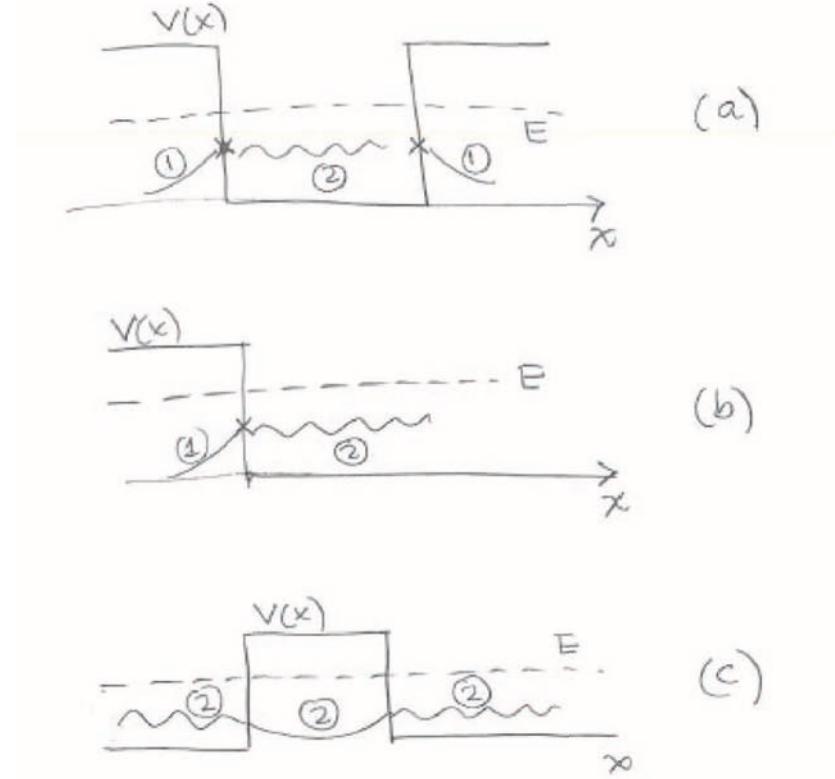


Figure 1: Discussing the number of constants needed to specify a solution. (a) Energy is smaller than the potential for $x \rightarrow \pm\infty$. (b) Energy is smaller than the potential for $x \rightarrow -\infty$ and larger than the potential for $x \rightarrow \infty$. (c) Energy is larger than the potential for $x \rightarrow \pm\infty$.

freedom is accounted in. We also have two boundary conditions at the interface. So we can expect a solution. Indeed there should be a solution for each value of the energy. The spectrum here is continuous and non-degenerate.

- (c) Two constants are needed here in each of the three regions: they multiply sines and cosines to the left and to the right, and multiply the two exponentials in the middle. Thus six constants and due to scaling just five unknowns. We still have four boundary conditions so there should be solutions. In fact, there are two solutions for each energy. We can understand this as follows. Think of using just one coefficient to the far left, say the coefficient multiplying the sine function. With one less coefficient we have the same number of unknowns as constraints so we should get one solution (for any E). We get another solution if we use the cosine function to the far left. So we have two solutions for each energy. The spectrum is continuous and doubly degenerate.

Figure 2 illustrates the spectrum of the Hamiltonian for a rather generic type of potential. Here V_+ is the top asymptote of the potential, V_- is the bottom asymptote of the potential, and V_0 is the lowest value of the potential. In the figure we indicate the type of spectrum for energies in the various intervals defined: $E > V_+$, then $V_- < E < V_+$, then $V_0 < E < V_-$ and finally $E < V_0$.

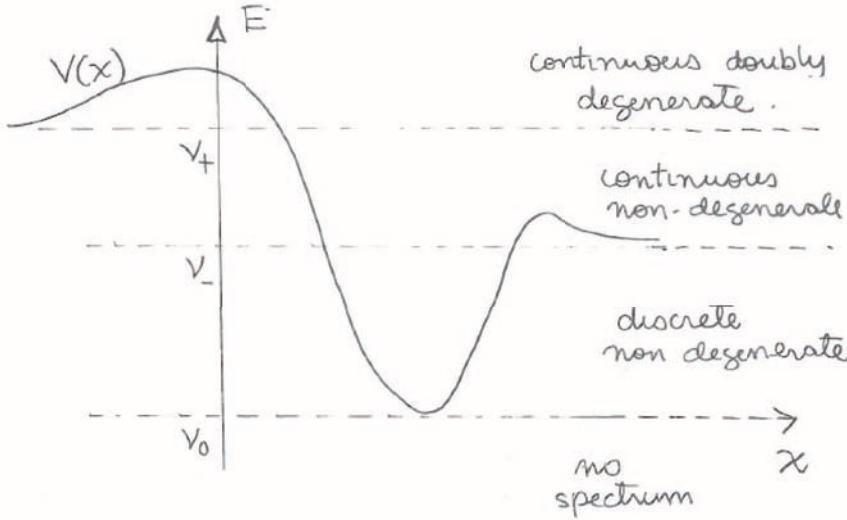


Figure 2: A generic potential and the type of spectrum for various energy ranges.

A **node** in a wavefunction is a point x_0 where $\psi(x_0) = 0$ (a zero of ψ) and $\psi'(x_0) \neq 0$. For a bound state the zeroes are nodes or the points at infinity (where typically $\psi' \rightarrow 0$).

Theorem 4 For the discrete bound-state spectrum of a one-dimensional potential let the allowed energies be $E_1 < E_2 < E_3 < \dots$ with E_1 the ground state energy. Let the associated energy eigenstates be $\psi_1, \psi_2, \psi_3, \dots$. The wavefunction ψ_1 has no nodes, ψ_2 has one node, and each consecutive wavefunction has one additional node. In conclusion ψ_n has $n - 1$ nodes.

We will not prove this theorem here. In fact you will show in the homework that ψ_{k+1} has at least one node between two consecutive zeroes of ψ_k . This implies that ψ_{k+1} has at least one more node than ψ_k . This can be illustrated in Figure 3 that shows a bound state $\psi_4(x)$ with three nodes at x_1, x_2 , and x_3 and zeroes at $x = -\infty$ and $x = \infty$. For ψ_5 there must be a node w_1 in $(-\infty, x_1]$, a node $w_2 \in (x_1, x_2)$ and so on until a last node $w_4 \in (x_3, \infty)$.

Example: Potential with five delta functions. We will discuss the bound states of the Schrödinger equation with potential

$$V(x) = -V_0 a \sum_{n=-2}^2 \delta(x - na). \quad (4.3)$$

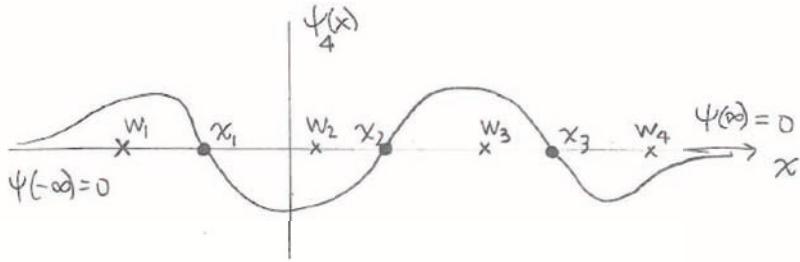


Figure 3: A wavefunction ψ_4 with three nodes (x_1, x_2, x_3) and zeroes at $x \pm \infty$. The next wavefunction ψ_5 must have four nodes, with positions indicated by w_1, w_2, w_3 and w_4 .

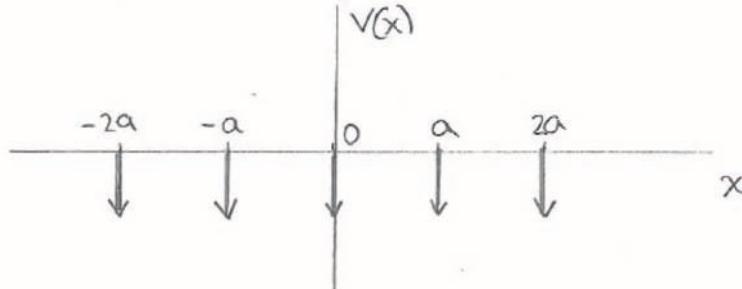


Figure 4: A potential $V(x)$ with five downwards pointing delta-functions.

This potential has delta functions at x equal to $-2a, -a, 0, a$, and $2a$, as shown in Figure 4.

We first examine the effect of the delta functions on the eigenstates. We will see that they produce discontinuities in ψ' at the position of the delta functions. We begin with the Schrödinger equation

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + V(x)\psi(x) = E\psi(x), \quad (4.4)$$

and integrate this equation from $a - \epsilon$ to $a + \epsilon$, where ϵ is a small value that we will take down to zero. By doing this we will get one out of the five delta functions to fire. We find

$$-\frac{\hbar^2}{2m} \int_{a-\epsilon}^{a+\epsilon} dx \frac{d^2\psi}{dx^2} + \int_{a-\epsilon}^{a+\epsilon} dx V(x)\psi(x) = E \int_{a-\epsilon}^{a+\epsilon} dx \psi(x). \quad (4.5)$$

The first term involves a total derivative, the second term just picks up the delta function at $x = a$, and the right hand side is evaluated by noting that since ψ is continuous its value at $x = a$ gives the leading contribution:

$$-\frac{\hbar^2}{2m} \frac{d\psi}{dx} \Big|_{a-\epsilon}^{a+\epsilon} - V_0 a \int_{a-\epsilon}^{a+\epsilon} dx \delta(x - a)\psi(x) = E(2\epsilon)\psi(a) + \mathcal{O}(\epsilon^2). \quad (4.6)$$

In the limit as $\epsilon \rightarrow 0$ we will denote $a + \epsilon$ as a^+ and $a - \epsilon$ as a^- . These labels are needed since ψ' has to be discontinuous at x . Indeed, we get

$$-\frac{\hbar^2}{2m} (\psi'(a^+) - \psi'(a^-)) - V_0 a \psi(a) = 0. \quad (4.7)$$

This implies that the discontinuity $\Delta\psi'$ of ψ' is given by

$$\Delta\psi'(a) \equiv \psi'(a^+) - \psi'(a^-) = \frac{2m}{\hbar^2} (-V_0 a) \psi(a).$$

The discontinuity of ψ' at the position of the delta function is proportional to the value of ψ at this point. The constant of proportionality is linear on the strength $V_0 a$ of the delta function. It follows that if the delta function of the potential is at a point where ψ vanishes then both ψ and ψ' are continuous and the delta function has no effect.

Let us now focus on bound states. These will be states with $E < 0$. The Schrödinger equation away from the delta functions is just

$$\psi'' = -\frac{2mE}{\hbar^2} \psi = \kappa^2 \psi, \text{ with } \kappa^2 \equiv -\frac{2mE}{\hbar^2} > 0, \kappa > 0. \quad (4.9)$$

The solutions are therefore the linear combinations

$$\psi(x) = ae^{-\kappa x} + be^{\kappa x}, \quad (4.10)$$

with a and b real constants to be determined (recall the wavefunction can be taken to be real). In Figure 5 we show these functions for $a > b > 0$. Note the the curves intersect just once. It follows that the wavefunction will never have a zero if a and b have the same sign and it will have exactly one zero if a and b have opposite signs.

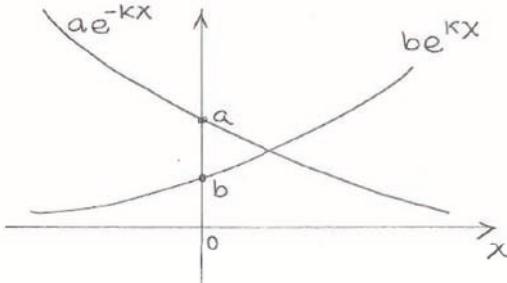


Figure 5: Plots of $ae^{-\kappa x}$ and $be^{\kappa x}$ with $a, b > 0$. This can be used to show that any linear superposition of these two functions can at most have one zero.

Let us then make the following remarks:

1. There cannot be zeroes of the wavefunction for $x \geq 2a$ (nor for $x \leq -2a$). For $x \geq 2a$ the solution, if non vanishing, must be of the form $ce^{-\kappa x}$. This can only have a zero if $c = 0$. In this case the wavefunction would vanish identically for $x \geq 2a$. This does not look good. Since $\psi(2a) = 0$ then ψ' is not discontinuous and, by continuity, a bit to the left of $2a$ both ψ and ψ' vanish. This is enough to make the solution vanish over the next interval $x \in (a, 2a)$. Continuing in this way we find that the solution for ψ would have to be zero everywhere. This is not acceptable.
2. There is at most one zero in between each pair of contiguous δ -functions. This follows because the solution must take the form (4.10) and we argued that such function can at most have one zero.
3. Zeroes appear at $x = 0$ for all the antisymmetric bound states. In those cases, there cannot be another zero in the interval $[-a, a]$. Zeroes may appear at $x = \pm a$, but this is presumably not generic. There are at most five bound states because the maximum number of nodes is four; one in between each delta function. All these five bound states exist if the delta functions are strong enough. The ground state is even, has no nodes and presumably looks like the one drawn in Figure 6.

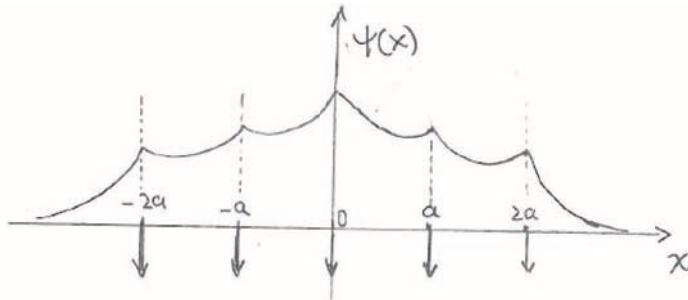


Figure 6: A sketch of the ground state wavefunction.

Exercise. Sketch the expected shapes of the four excited bound states of the potential.

5 Variational Principle

Consider a system with Hamiltonian \hat{H} and focus on the time-independent Schrödinger equation:

$$\hat{H}\psi(\vec{x}) = E\psi(\vec{x}). \quad (5.11)$$

Let us assume that the system is such that it has a collection of energy eigenstates that are normalizable. This collection includes a ground state with ground state energy E_{gs} . Note the

use of \vec{x} : our discussion applies to quantum systems in any number of spatial dimensions. Our first goal is to learn something about the ground state energy without solving the Schrödinger equation nor trying to figure out the ground state wavefunction.

For this purpose, consider an *arbitrary* normalized wavefunction $\psi(\vec{x})$:

$$\int d\vec{x} \psi^*(\vec{x}) \psi(\vec{x}) = 1. \quad (5.12)$$

By arbitrary we mean a wavefunction that need not satisfy the time-independent Schrödinger equation, a wavefunction that need not be an energy eigenstate. Then we claim the ground state energy E_{gs} of the Hamiltonian is smaller or equal than the expectation value of \hat{H} in this arbitrary normalized ψ , namely,

$$E_{gs} \leq \langle \hat{H} \rangle_\psi = \int d\vec{x} \psi^*(\vec{x}) \hat{H} \psi(\vec{x}), \quad \text{Normalized } \psi. \quad (5.13)$$

The wavefunction $\psi(\vec{x})$ here is sometimes called a *trial* wavefunction. When the right-hand side of the above inequality is evaluated we get an energy and learn that the ground state energy must be smaller or equal to the value we get. Thus any trial wavefunction provides an *upper bound* for the ground state energy. Better and better trial wavefunctions will produce lower and lower upper bounds. Note that if the trial wavefunction was set equal to the (unknown) ground-state wavefunction, the expectation value of \hat{H} becomes exactly E_{gs} and the inequality is saturated.

Let us prove (5.13). For simplicity, we will consider here the case where the energy eigenstates $\psi_n(\vec{x})$ of \hat{H} are denumerable and their corresponding energies E_n are ordered as

$$E_{gs} = E_1 \leq E_2 \leq E_3 \leq \dots. \quad (5.14)$$

Of course $\hat{H}\psi_n = E_n\psi_n$. Since the energy eigenstates are complete, any trial wavefunction can be expanded in terms of them (see (2.18)):

$$\psi(\vec{x}) = \sum_{n=1}^{\infty} b_n \psi_n(\vec{x}). \quad (5.15)$$

Such a ψ is not an energy eigenstate in general. The normalization condition (5.12) gives us,

$$\sum_{n=1}^{\infty} |b_n|^2 = 1. \quad (5.16)$$

The evaluation of the right-hand side in (5.13) was done before in (2.28) so we have

$$\int d\vec{x} \psi^*(\vec{x}) \hat{H} \psi(\vec{x}) = \sum_{n=1}^{\infty} |b_n|^2 E_n. \quad (5.17)$$

Since $E_n \geq E_1$ for all n , we can replace the E_n on the above right-hand side for E_1 getting a smaller or equal value:

$$\int d\vec{x} \psi^*(\vec{x}) \hat{H} \psi(\vec{x}) \geq \sum_{n=1}^{\infty} |b_n|^2 E_1 = E_1 \sum_{n=1}^{\infty} |b_n|^2 = E_1 = E_{gs}, \quad (5.18)$$

where we used (5.16). This is in fact the claim in (5.13).

It is sometimes more convenient not to worry about the normalization of the trial wavefunctions. Given a trial wavefunction ψ that is not normalized, the wavefunction

$$\frac{\psi(x)}{\sqrt{N}} \quad \text{with } N = \int d\vec{x} \psi^*(\vec{x}) \psi(\vec{x}), \quad (5.19)$$

is normalized and can be used in (5.13). We therefore find that

$$E_{gs} \leq \frac{\int d\vec{x} \psi^*(\vec{x}) \hat{H} \psi(\vec{x})}{\int d\vec{x} \psi^*(\vec{x}) \psi(\vec{x})} \equiv \mathcal{F}[\psi]. \quad (5.20)$$

This formula can be used for trial wavefunctions that are not normalized. We also introduced the definition of the functional $\mathcal{F}[\psi]$. A functional is a machine that given a function gives us a number. Our result states that the ground state energy arises as the minimum value that the functional can take.

One application of this variational principle is to find good upper bounds for the ground state energy of quantum systems that are not exactly solvable. For this purpose it is useful to construct trial wavefunctions $\psi(\vec{x}; \beta_1, \beta_2, \dots, \beta_m)$ that depend on a set of parameters β . One then computes the expectation value $\langle \hat{H} \rangle_\psi$ which, of course, is a function of the parameters. Any random values for the parameters will give an upper bound for the ground state energy, but by minimizing $\langle \hat{H} \rangle_\psi$ over the parameter space we get the lowest possible upper bound consistent with the chosen form for the trial wavefunction.

Example. (Griffiths). Consider a one-dimensional problem with the delta function potential

$$V(x) = -\alpha \delta(x), \quad \alpha > 0. \quad (5.21)$$

In this problem the ground state energy is calculable exactly and one has

$$E_{gs} = -\frac{m\alpha^2}{2\hbar^2}. \quad (5.22)$$

So this problem is just for illustration. Consider an unnormalized gaussian trial wavefunction, with a real parameter β :

$$\psi(x) = e^{-\frac{1}{2}\beta^2 x^2}, \quad \int_{-\infty}^{\infty} dx \psi^2 = \frac{\sqrt{\pi}}{\beta}. \quad (5.23)$$

The functional \mathcal{F} in (5.20) is then¹

$$\begin{aligned}
\frac{\int dx \psi^*(x) \hat{H} \psi(x)}{\int dx \psi^*(x) \psi(x)} &= \frac{\beta}{\sqrt{\pi}} \int dx e^{-\frac{1}{2}\beta^2 x^2} \left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} - \alpha \delta(x) \right) e^{-\frac{1}{2}\beta^2 x^2} \\
&= \frac{\beta}{\sqrt{\pi}} \frac{\hbar^2}{2m} \int dx \left[\frac{d}{dx} e^{-\frac{1}{2}\beta^2 x^2} \right]^2 - \frac{\beta}{\sqrt{\pi}} \alpha \\
&= \frac{\beta}{\sqrt{\pi}} \frac{\hbar^2}{2m} \frac{\beta \sqrt{\pi}}{2} - \frac{\beta}{\sqrt{\pi}} \alpha \\
&= \frac{\beta^2 \hbar^2}{4m} - \frac{\beta}{\sqrt{\pi}} \alpha.
\end{aligned} \tag{5.24}$$

The first term on the last right-hand side is the kinetic energy and the second term is the potential energy. For any value of β the final expression above provides an upper bound for the ground state energy, and the best upper bound is the lowest one. We thus have that the ground state energy satisfies

$$E_{gs} \leq \text{Min}_\beta \left(\frac{\beta^2 \hbar^2}{4m} - \frac{\beta}{\sqrt{\pi}} \alpha \right). \tag{5.25}$$

The minimum is easily found

$$\beta = \frac{2m\alpha}{\hbar^2 \sqrt{\pi}} \quad \rightarrow \quad E_{gs} \leq -\frac{m\alpha^2}{\pi \hbar^2} = \frac{2}{\pi} \left(-\frac{m\alpha^2}{2\hbar^2} \right). \tag{5.26}$$

Comparing with (5.22) we see that the bound we found is in fact $\frac{2}{\pi} E_{gs} \simeq 0.64 E_{gs}$. The trial wavefunction brought us to about 64% of the correct value.

In the exercises you will develop the following results:

1. With trial wavefunctions orthogonal to the ground state, the functional \mathcal{F} gives upper bounds for the energy of the first excited state.
2. For any attractive one-dimensional potential (a nowhere positive potential that approaches zero at infinity) there is a bound state, namely, a state with energy less than zero.
3. We have shown that the functional $\mathcal{F}[\psi]$ has a minimum for ψ equal to the ground state wavefunction. Interestingly, this functional is stationary at each and every energy eigenstate. For eigenstates of energies higher than the ground state \mathcal{F} has a saddle point.

¹We use the integrals $\int du e^{-u^2} = \sqrt{\pi}$ and $\int du u^2 e^{-u^2} = \frac{1}{2}\sqrt{\pi}$.

6 Position and momentum

In quantum mechanics the position operator \hat{x} and the momentum operator \hat{p} do not commute. They satisfy the commutation relation

$$[\hat{x}, \hat{p}] = i\hbar. \quad (6.27)$$

When we deal with wavefunctions $\psi(x)$ the position operator acts on them in a simple way.² We define

$$\hat{x}\psi(x) \equiv x\psi(x). \quad (6.28)$$

In words the position operator acting on an x -dependent wavefunction simply multiplies the wavefunction by x . In quantum mechanics it is useful think of states as vectors and operators as matrices. A wavefunction for a particle on the box $0 \leq x \leq a$, for example can be thought as vector with many components, each one giving the value of the function at a specific point. To make this concrete one discretizes the space into small intervals of size ϵ such that $N\epsilon = a$. In that case we can represent the information in $\psi(x)$ in a large column vector

$$\psi(x) \longleftrightarrow \begin{pmatrix} \psi(0) \\ \psi(\epsilon) \\ \psi(2\epsilon) \\ \vdots \\ \psi(N\epsilon) \end{pmatrix}. \quad (6.29)$$

The $N + 1$ component column vector summarizes the values of the wavefunction at equally separated points. N is some kind of regulator: a precise description requires $N \rightarrow \infty$ or $\epsilon \rightarrow 0$. Associated with the description (6.29) the operator \hat{x} can be viewed as the $(N + 1) \times (N + 1)$ diagonal matrix

$$\hat{x} \longleftrightarrow \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \epsilon & 0 & \dots & 0 \\ 0 & 0 & 2\epsilon & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & N\epsilon \end{pmatrix}. \quad (6.30)$$

You can see that the action of the matrix \hat{x} on the vector (6.29) gives the vector

$$\begin{pmatrix} 0 \cdot \psi(0) \\ \epsilon \cdot \psi(\epsilon) \\ 2\epsilon \cdot \psi(2\epsilon) \\ \vdots \\ N\epsilon \cdot \psi(N\epsilon) \end{pmatrix}, \quad (6.31)$$

²The time dependence is irrelevant to the present discussion, which applies without changes to time-dependent wavefunctions $\Psi(x, t)$.

which is indeed the representation of $x\psi(x)$. Given our definition of the action of \hat{x} , expectation values in normalized states are naturally defined by

$$\langle \hat{x} \rangle \equiv \int dx' \psi^*(x') (x' \psi(x')) . \quad (6.32)$$

Are there eigenstates of the \hat{x} operator? Yes, but their are not normalizable. An eigenstate of \hat{x} must be a localized state, and the obvious candidate is a delta function. Defining

$$\psi_{x_0}(x) \equiv \delta(x - x_0) , \quad (6.33)$$

we verify that

$$\hat{x}\psi_{x_0}(x) = x\psi_{x_0}(x) = x\delta(x - x_0) = x_0\delta(x - x_0) = x_0\psi_{x_0}(x) , \quad (6.34)$$

confirming that $\psi_{x_0}(x)$ is an eigenstate of \hat{x} with eigenvalue x_0 . A delta function cannot be normalized, so the position eigenstates are not normalizable.

When we speak about the state of a particle and describe it with the wavefunction $\psi(x)$ we are using the “position” representation of the state. The momentum operator in the position representation is given by

$$\hat{p} \equiv \frac{\hbar}{i} \frac{\partial}{\partial x} \quad (6.35)$$

This is to say that acting on a wavefunction we have

$$\hat{p}\psi(x) = \frac{\hbar}{i} \frac{d\psi}{dx} . \quad (6.36)$$

Note that the commutation relation (6.27) is satisfied by the above definitions, as we can check acting on any wavefunction:

$$\begin{aligned} [\hat{x}, \hat{p}]\psi(x) &= (\hat{x}\hat{p} - \hat{p}\hat{x})\psi(x) \\ &= \hat{x}\hat{p}\psi(x) - \hat{p}\hat{x}\psi(x) \\ &= \hat{x} \frac{\hbar}{i} \frac{d\psi}{dx} - \hat{p}x\psi(x) \\ &= \frac{\hbar}{i} x \frac{d\psi}{dx} - \frac{\hbar}{i} \frac{d}{dx}(x\psi) \\ &= \frac{\hbar}{i} x \frac{d\psi}{dx} - \frac{\hbar}{i}\psi - \frac{\hbar}{i}x \frac{d\psi}{dx} \\ &= i\hbar\psi(x) . \end{aligned} \quad (6.37)$$

Since the wavefunction is arbitrary, we have verified that our explicit representation of the operators \hat{x} (by multiplication) and \hat{p} (by differentiation) satisfies the commutation relation $[\hat{x}, \hat{p}] = i\hbar$.

Eigenstates of the momentum operator exist and are not normalizable. Defining

$$\psi_p(x) \equiv \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}, \quad (6.38)$$

we readily confirm that

$$\hat{p}\psi_p(x) = \frac{\hbar}{i} \frac{\partial}{\partial x} \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} = p \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} = p\psi_p(x). \quad (6.39)$$

So $\psi_p(x)$ is a momentum eigenstate with momentum eigenvalue p . It is a plane wave.

The so-called momentum representation is mathematically described by Fourier transforms. The Fourier transform $\tilde{\psi}(p)$ of $\psi(x)$ is defined by

$$\tilde{\psi}(p) \equiv \int_{-\infty}^{\infty} dx \frac{e^{-ipx/\hbar}}{\sqrt{2\pi\hbar}} \psi(x). \quad (6.40)$$

The function $\tilde{\psi}(p)$ encodes the same amount of information as $\psi(x)$. We call $\tilde{\psi}(p)$ the momentum space representation of the state. Clearly for each value of p , $\tilde{\psi}$ is a linear superposition of values of $\psi(x)$ for all x . We can view the Fourier transformation as a linear transformation, the action of a matrix that depends on p on the vector that represents $\psi(x)$. The inverse Fourier transform is written as

$$\psi(x) = \int_{-\infty}^{\infty} dp \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} \tilde{\psi}(p). \quad (6.41)$$

We can view this formula as an expansion of $\psi(x)$ in a basis of momentum eigenstates, with $\tilde{\psi}(p)$ the expansion coefficients.

We have seen that $\psi(x)$ and $\tilde{\psi}(p)$ are just two different representations of the same state:

$$\psi(x) \longleftrightarrow \tilde{\psi}(p). \quad (6.42)$$

The arrow above is implemented by Fourier Transformation. Calculate now the action of $\frac{\hbar}{i} \frac{d}{dx}$ on (6.41)

$$\frac{\hbar}{i} \frac{d}{dx} \psi(x) = \frac{\hbar}{i} \frac{d}{dx} \int_{-\infty}^{\infty} dp \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} \tilde{\psi}(p) = \int_{-\infty}^{\infty} dp \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} p \tilde{\psi}(p). \quad (6.43)$$

In the language of (6.42) we write this as

$$\frac{\hbar}{i} \frac{d}{dx} \psi(x) \longleftrightarrow p \tilde{\psi}(p). \quad (6.44)$$

We see that the momentum operator, viewed as the action of $\frac{\hbar}{i} \frac{d}{dx}$ in coordinate space, is simply multiplication by p on momentum space wavefunctions $\tilde{\psi}$:

$$\hat{p} \tilde{\psi}(p) = p \tilde{\psi}(p). \quad (6.45)$$

This is, of course, perfectly analogous to the way that \hat{x} acts on position space wavefunctions.

Exercise. Verify that acting on momentum space wavefunctions the \hat{x} operator is represented by

$$\hat{x} \equiv i\hbar \frac{d}{dp}, \quad (\text{momentum representation}) \quad (6.46)$$

You can do this in two ways. Working with Fourier transforms, or by verifying (as in (6.37)) that it is consistent with $[\hat{x}, \hat{p}] = i\hbar$ acting on momentum space wavefunctions.

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

TWO STATE SYSTEMS

B. Zwiebach
November 15, 2013

Contents

1	Introduction	1
2	Spin precession in a magnetic field	2
3	The general two-state system viewed as a spin system	5
4	The ammonia molecule as a two-state system	7
5	Ammonia molecule in an electric field	11
6	Nuclear Magnetic Resonance	17

1 Introduction

A two-state system does not just have two states! It has two *basis* states, namely the state space is a two-dimensional complex vector space. For such a state space the Hamiltonian can be viewed as the most general Hermitian 2×2 matrix. In the case when the Hamiltonian is time-independent, this Hermitian matrix is characterized by four real numbers.

Two-state systems are idealizations that are valid when other degrees of freedom are ignored. A spin one-half particle is a two-state system with regards to spin, but being a particle, it may move and thus has position, or momentum degrees of freedom that imply a much larger, higher dimensional state space. Only if we ignore these degrees of freedom – perhaps because the particle is at rest – we can speak of a two-state system.

We will study here two specific examples of two-state systems. The first will be the ammonia molecule, which exhibits curious oscillations. The second will be spin one-half particles as used in nuclear magnetic resonance.

The mathematics of two-state systems is always the same. In fact any two-state system can be visualized as a spin system, and this will sometimes be quite useful.

2 Spin precession in a magnetic field

Let us first recall our earlier discussion of magnetic dipole moments. Classically we had the following relation valid for a charged particle

$$\boldsymbol{\mu} = \frac{q}{2m} \mathbf{S} \quad (2.1)$$

where $\boldsymbol{\mu}$ is the dipole moment, q is the charge of the particle, m its mass, and \mathbf{S} is its angular momentum, arising from its spinning. In the quantum world this equation gets modified by a constant unit-free factor g , different for each particle:

$$\boldsymbol{\mu} = g \frac{q}{2m} \mathbf{S} = g \frac{q\hbar}{2m} \frac{\mathbf{S}}{\hbar}. \quad (2.2)$$

Here $q\hbar/(2m)$ has the units of dipole moment. If we consider electrons and protons the following definitions are thus natural:

$$\begin{aligned} \text{Born magneton: } \mu_B &= \frac{e\hbar}{2m_e} = 5.78 \times 10^{-11} \frac{\text{MeV}}{\text{Tesla}}, \\ \text{Nuclear magneton: } \mu_N &= \frac{e\hbar}{2m_p} = 3.15 \times 10^{-14} \frac{\text{MeV}}{\text{Tesla}}. \end{aligned} \quad (2.3)$$

Note that the nuclear magneton is about two-thousand times smaller than the Bohr magneton. Nuclear magnetic dipole moments are much smaller than that of the electron! Including the g constant we have the following results. For an electron $g = 2$ and since the electron charge is negative we get

$$\boldsymbol{\mu}_e = -2 \mu_B \frac{\mathbf{S}}{\hbar}. \quad (2.4)$$

The dipole moment and the angular momentum are antiparallel. For a proton, the experimental result is

$$\boldsymbol{\mu}_p = 2.79 \mu_N \frac{\mathbf{S}}{\hbar}. \quad (2.5)$$

The neutron is neutral, so one would expect no magnetic dipole moment. But the neutron is not elementary: it is made by electrically charged quarks. A dipole moment is thus possible, depending on the way quarks are distributed. Indeed, experimentally,

$$\boldsymbol{\mu}_n = -1.91 \mu_N \frac{\mathbf{S}}{\hbar}. \quad (2.6)$$

Somehow the negative charge beats the positive charge in its contribution to the dipole moment of the neutron.

For notational convenience we introduce the constant γ from

$$\boldsymbol{\mu} = \gamma \mathbf{S}, \quad \text{with } \gamma = \frac{gq}{2m}.$$

(2.7)

If we insert the particle in a magnetic field \mathbf{B} , the Hamiltonian H_S for the spin system is

$$H_S = -\boldsymbol{\mu} \cdot \mathbf{B} = -\gamma \mathbf{B} \cdot \mathbf{S} = -\gamma (B_x \hat{S}_x + B_y \hat{S}_y + B_z \hat{S}_z). \quad (2.8)$$

If, for example we have a magnetic field $\vec{B} = B\hat{z}$ along the z axis, the Hamiltonian is

$$H_S = -\gamma B \hat{S}_z \quad (2.9)$$

The associated time evolution unitary operator is

$$\mathcal{U}(t, 0) = \exp\left(-\frac{iH_S t}{\hbar}\right) = \exp\left(-\frac{i(-\gamma B t)\hat{S}_z}{\hbar}\right) \quad (2.10)$$

We now recall a result that was motivated in the homework. You examined a unitary operator $R_{\mathbf{n}}(\alpha)$ defined by a unit vector \mathbf{n} and an angle α , and given by

$$R_{\mathbf{n}}(\alpha) = \exp\left(-\frac{i\alpha \hat{S}_{\mathbf{n}}}{\hbar}\right), \quad \text{with } \hat{S}_{\mathbf{n}} \equiv \mathbf{n} \cdot \mathbf{S}. \quad (2.11)$$

You found evidence that when acting on a spin state, this operator rotates it by an angle α about the axis defined by the vector \mathbf{n} . If we now compare (2.11) and (2.10) we conclude that $\mathcal{U}(t, 0)$ should generate a rotation by the angle $(-\gamma B t)$ about the z -axis. We now confirm this explicitly.

Consider a spin pointing at time equal zero along the direction specified by the angles (θ_0, ϕ_0) :

$$|\Psi, 0\rangle = \cos \frac{\theta_0}{2} |+\rangle + \sin \frac{\theta_0}{2} e^{i\phi_0} |-\rangle \quad (2.12)$$

Given the Hamiltonian $H_S = -\gamma B \hat{S}_z$ in (2.9) we have

$$H_S |\pm\rangle = \mp \frac{\gamma B \hbar}{2} \quad (2.13)$$

Then we have

$$\begin{aligned} |\Psi, t\rangle &= e^{-iH_S t/\hbar} |\Psi, 0\rangle = e^{-iH_S t/\hbar} \left(\cos \frac{\theta_0}{2} |+\rangle + \sin \frac{\theta_0}{2} e^{i\phi_0} |-\rangle \right) \\ &= \cos \frac{\theta_0}{2} e^{-iH_S t/\hbar} |+\rangle + \sin \frac{\theta_0}{2} e^{i\phi_0} e^{-iH_S t/\hbar} |-\rangle \\ &= \cos \frac{\theta_0}{2} e^{+i\gamma B t/2} |+\rangle + \sin \frac{\theta_0}{2} e^{i\phi_0} e^{-i\gamma B t/2} |-\rangle \end{aligned} \quad (2.14)$$

using (2.13). To recognize the resulting state it is convenient to factor out the phase that multiplies the $|+\rangle$ state:

$$|\Psi, t\rangle = e^{+i\gamma B t/2} \left(\cos \frac{\theta_0}{2} |+\rangle + \sin \frac{\theta_0}{2} e^{i(\phi_0 - \gamma B t)} |-\rangle \right). \quad (2.15)$$

Since the overall phase is not relevant we can now recognize the spin state as the state corresponding to the vector $\vec{n}(t)$ defined by angles

$$\begin{aligned}\theta(t) &= \theta_0, \\ \phi(t) &= \phi_0 - \gamma B t.\end{aligned}\tag{2.16}$$

Keeping θ constant while changing ϕ indeed corresponds to a rotation about the z axis and, after time t , the spin has rotated an angle $(-\gamma B t)$ as claimed above.

In fact spin states in a magnetic field precess in exactly the same way that magnetic dipoles in classical electromagnetism precess. The main fact from electromagnetic theory that we need is that in a magnetic field a dipole moment experiences a torque $\boldsymbol{\tau}$ given by

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}.\tag{2.17}$$

Then the familiar mechanics equation for the rate of change of angular momentum being equal to the torque gives

$$\frac{d\mathbf{S}}{dt} = \boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B} = \gamma \mathbf{S} \times \mathbf{B},\tag{2.18}$$

which we write as

$$\frac{d\mathbf{S}}{dt} = -\gamma \mathbf{B} \times \mathbf{S}.\tag{2.19}$$

We recognize that this equation states that the time dependent vector is rotating with angular velocity $\vec{\omega}_L$ given by

$$\boxed{\boldsymbol{\omega}_L = -\gamma \mathbf{B}.}\tag{2.20}$$

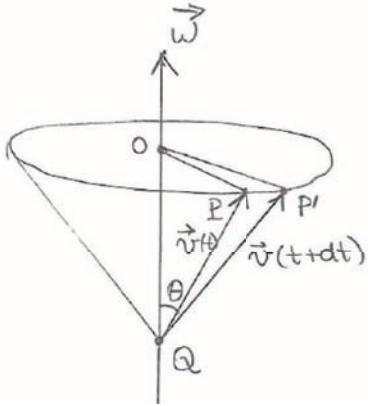
This is the so-called Larmor frequency. Indeed, this identification is standard in mechanics. A vector \mathbf{v} rotating with angular velocity $\boldsymbol{\omega}$ satisfies the differential equation

$$\frac{d\mathbf{v}}{dt} = \boldsymbol{\omega} \times \mathbf{v}.\tag{2.21}$$

You can convince yourself of this with the help of a simple picture (see Figure 1). Also note that the differential equation shows that the derivative of \mathbf{v} , given by the right-hand side, is orthogonal to \mathbf{v} because the cross product involves \mathbf{v} . This is as it should when the vector \mathbf{v} is rotated. Indeed, show that the above differential equation implies that $\frac{d}{dt}\mathbf{v} \cdot \mathbf{v} = 0$, so that the length of \mathbf{v} is unchanged.

The Hamiltonian of a general spin in a magnetic field (2.8) is then

$$\boxed{H_S = -\boldsymbol{\mu} \cdot \mathbf{B} = -\gamma \mathbf{B} \cdot \mathbf{S} = \boldsymbol{\omega}_L \cdot \mathbf{S}.}\tag{2.22}$$



$$\begin{aligned}
 \angle POP' &= \omega dt \\
 |OP| = |OP'| &= v \sin \theta \\
 |PP'| &= (\omega dt) v \sin \theta \\
 \vec{PP}' &= \vec{\omega} \times \vec{v} dt
 \end{aligned}$$

Figure 1: The vector $\mathbf{v}(t)$ and an instant later the vector $\mathbf{v}(t + dt)$. The angular velocity vector $\boldsymbol{\omega}$ is along the axis and \mathbf{v} rotates about is origin Q . At all times the vector \mathbf{v} and $\boldsymbol{\omega}$ make an angle θ . The calculations to the right should convince you that (2.21) is correct.

For time independent magnetic fields $\boldsymbol{\omega}_L$ is also time independent and the evolution operator is

$$\mathcal{U}(t, 0) = \exp(-iH_S t/\hbar) = \exp\left(-i \frac{\boldsymbol{\omega}_L \cdot \mathbf{S}}{\hbar} t\right). \quad (2.23)$$

If we write

$$\boldsymbol{\omega}_L = \omega_L \mathbf{n}, \quad \mathbf{n} \cdot \mathbf{n} = 1, \quad (2.24)$$

we have

$$\mathcal{U}(t, 0) = \exp\left(-i \frac{\omega_L t \hat{S}_{\mathbf{n}}}{\hbar}\right) = R_{\mathbf{n}}(\omega_L t), \quad (2.25)$$

where we compared with (2.11). The time evolution operator $\mathcal{U}(t, 0)$ rotates the spin states by the angle $\omega_L t$ about the \mathbf{n} axis. In other words

With $H_S = \boldsymbol{\omega}_L \cdot \mathbf{S}$ spin states precess with angular velocity $\boldsymbol{\omega}_L$.

3 The general two-state system viewed as a spin system

The most general time-independent Hamiltonian for a two-state system is a hermitian operator represented by the most general hermitian two-by-two matrix H . In doing so we are using some orthonormal basis $\{|1\rangle, |2\rangle\}$. In any such basis the matrix can be characterized by four real constants $g_0, g_1, g_2, g_3 \in \mathbb{R}$ as follows:

$$H = \begin{pmatrix} g_0 + g_3 & g_1 - ig_2 \\ g_1 + ig_2 & g_0 - g_3 \end{pmatrix} = g_0 \mathbf{1} + g_1 \sigma_1 + g_2 \sigma_2 + g_3 \sigma_3. \quad (3.27)$$

On the right-hand side we wrote the matrix as a sum of matrices, where $\mathbf{1}$ and the Pauli matrices σ_i , $i = 1, 2, 3$ are hermitian. We view (g_1, g_2, g_3) as a vector \mathbf{g} and then define

$$\mathbf{g} \cdot \boldsymbol{\sigma} \equiv g_1\sigma_2 + g_2\sigma_2 + g_3\sigma_3. \quad (3.28)$$

In this notation

$$H = g_0\mathbf{1} + \mathbf{g} \cdot \boldsymbol{\sigma}. \quad (3.29)$$

It is again convenient to introduce the magnitude g and the direction \mathbf{n} of \mathbf{g} :

$$\mathbf{g} = g\mathbf{n}, \quad \mathbf{n} \cdot \mathbf{n} = 1, \quad g = \sqrt{g_1^2 + g_2^2 + g_3^2}. \quad (3.30)$$

Now the Hamiltonian reads

$$H = g_0\mathbf{1} + g\mathbf{n} \cdot \boldsymbol{\sigma}. \quad (3.31)$$

Recall now that the spin states $|\mathbf{n}; \pm\rangle$ are eigenstates of $\mathbf{n} \cdot \boldsymbol{\sigma}$

$$\mathbf{n} \cdot \boldsymbol{\sigma} |\mathbf{n}; \pm\rangle = \pm |\mathbf{n}; \pm\rangle. \quad (3.32)$$

In writing the spin states, however, you must recall that what we call the z -up and z -down states are just the first and second basis states: $|+\rangle = |1\rangle$ and $|-\rangle = |2\rangle$. With this noted, the spin states $|\mathbf{n}; \pm\rangle$ are indeed the eigenstates of H , since using the last two equations above we have

$$H|\mathbf{n}; \pm\rangle = (g_0 \pm g)|\mathbf{n}; \pm\rangle. \quad (3.33)$$

This also shows that the energy eigenvalues are $g_0 \pm g$. In summary:

$$\text{Spectrum: } |\mathbf{n}; +\rangle \text{ with energy } g_0 + g, \quad |\mathbf{n}; -\rangle \text{ with energy } g_0 - g. \quad (3.34)$$

Thus our spin states allow us to write the general solution for the spectrum of the Hamiltonian (again, writing $|+\rangle = |1\rangle$ and $|-\rangle = |2\rangle$). Clearly the $|1\rangle$ and $|2\rangle$ states will generally have nothing to do with spin states. They are the basis states of any two-state system.

To understand the time evolution of states with the Hamiltonian H , we first rewrite H in terms of the spin operators, instead of the Pauli matrices, recalling that $\mathbf{S} = \frac{\hbar}{2}\boldsymbol{\sigma}$. Using (3.29) we find

$$H = g_0\mathbf{1} + \frac{2}{\hbar}\mathbf{g} \cdot \mathbf{S}. \quad (3.35)$$

Comparison with the spin Hamiltonian $H_S = \boldsymbol{\omega}_L \cdot \mathbf{S}$ shows that in the system described by H the states precess with angular velocity $\boldsymbol{\omega}$ given by

$$\boldsymbol{\omega} = \frac{2}{\hbar} \mathbf{g}. \quad (3.36)$$

Note that the part $g_0\mathbf{1}$ of the Hamiltonian H does not rotate states during time evolution; it simply multiplies states by the time-dependent phase $\exp(-ig_0t/\hbar)$.

Operationally, if H is known, the vector ω above is immediately calculable. And given a normalized state $\alpha|1\rangle + \beta|2\rangle$ of the system ($|\alpha|^2 + |\beta|^2 = 1$), we can identify the corresponding spin state $|\mathbf{n};+\rangle = \alpha|+\rangle + \beta|-\rangle$. The time evolution of the spin state is due to Larmor precession and is intuitively understood. With this result, the time evolution of the state in the original system is simply obtained by letting $|+\rangle \rightarrow |1\rangle$ and $|-\rangle \rightarrow |2\rangle$ in the precessing spin state.

4 The ammonia molecule as a two-state system

The ammonia molecule NH_3 is composed of four atoms, one nitrogen and three hydrogen. Ammonia is naturally a gas, without color, but with a pungent odor. It is mostly used for fertilizers, and also for cleaning products and pharmaceuticals.

The ammonia molecule takes the shape of a flattened tetrahedron. If we imagine the three hydrogen atoms forming an equilateral triangle at the base, the nitrogen atom sits atop. The angle formed between any two lines joining the nitrogen to the hydrogen is about 108° – this indeed corresponds to a flattened tetrahedron since a regular tetrahedron would have a 60° angle. If the nitrogen was pushed all the way down to the base, the angle would be 120° .

The ammonia molecule has electronic excitations, vibrational excitations and rotational excitations. Those must largely be ignored in the two-state description of the molecule. The two states arise from transitions in which the nitrogen atom flips from being above the fixed hydrogen plane to being below the hydrogen plane. Since such a flip could be mimicked by a full rotation of the molecule, we can describe the transition more physically by considering the molecule spinning about the axis perpendicular to the hydrogen plane, with the N up, where up is the direction of the angular momentum. The transition would have the N down, or against the angular momentum of the rotating molecule.

More briefly, the two states are: nitrogen up, or nitrogen down. Both are classically stable configurations separated by a potential energy barrier. In classical mechanics these are the two options and they are degenerate in energy.

As long as the energy barrier is not infinite, in quantum mechanics the degeneracy is broken. This, of course is familiar. We can roughly represent the potential experienced by the nitrogen atom as the potential $V(z)$ in figure 3, where the two equilibrium positions of the nitrogen are at $\pm z_0$ and they are separated by a large barrier. In such a potential the ground state, which is symmetric, and the first excited state, which is antisymmetric, are almost degenerate in energy when the barrier is high. If the potential barrier was infinite, the two possible eigenstates would be the nitrogen wavefunction localized about z_0 and the nitrogen wavefunction localized about $-z_0$. Moreover, those states would be degenerate in energy. But with a large but finite barrier the ground state is represented by a wavefunction $\psi_g(z)$ even in z , as shown below the potential.

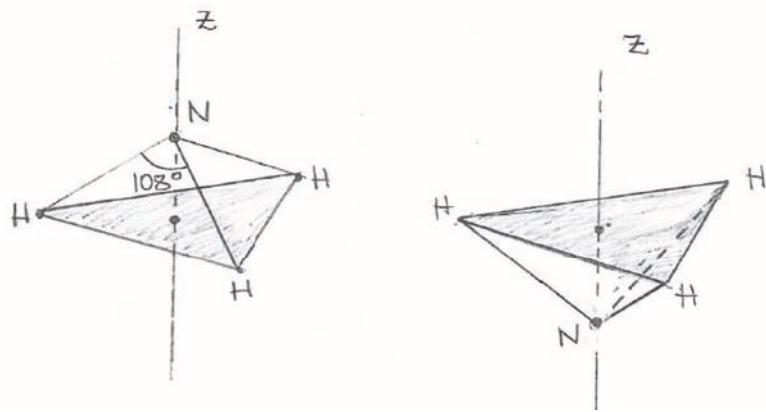


Figure 2: The ammonia molecule looks like a flattened tetrahedron. The nitrogen atom can be up or down with respect to the plane defined by the three hydrogen atoms. These are the two states of the ammonia molecule.

This even wavefunction is roughly the superposition, with the same sign, of the two localized wavefunctions. The next excited state $\psi_e(z)$ is odd in z and is roughly the superposition, this time with opposite signs, of the two localized wavefunctions.

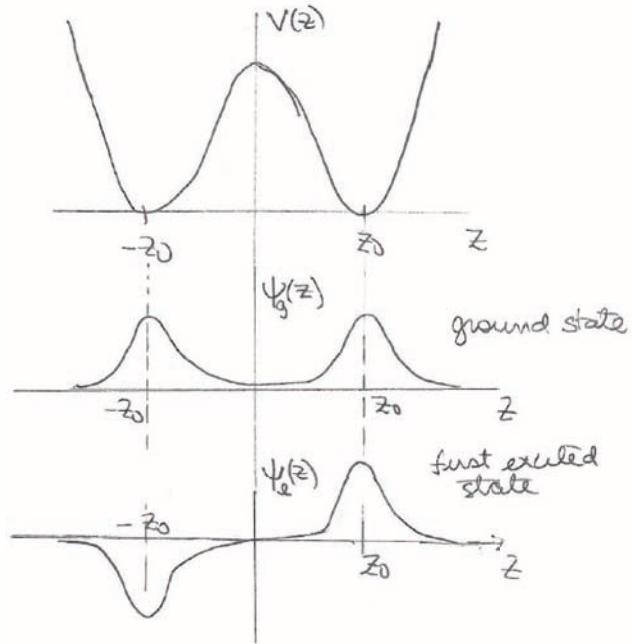


Figure 3: The potential $V(z)$ experienced by the nitrogen atom. There are two classically stable positions $\pm z_0$. The ground state and (first) excited state wavefunctions $\psi_g(z)$ and $\psi_e(z)$ are sketched below the potential.

Let us attempt a quantitative description of the situation. Let us label the two possible states:

$$|\uparrow\rangle \text{ is nitrogen up, } |\downarrow\rangle \text{ is nitrogen down} \quad (4.37)$$

We can associate to $|\uparrow\rangle$ a (positive) wavefunction localized around z_0 and to $|\downarrow\rangle$ a (positive) wavefunction localized around $-z_0$. Suppose the energy barrier is infinite. In this case the two states above must be energy eigenstates with the same energy E_0 :

$$\begin{aligned} H|\uparrow\rangle &= E_0|\uparrow\rangle, \\ H|\downarrow\rangle &= E_0|\downarrow\rangle. \end{aligned} \quad (4.38)$$

The energy E_0 is arbitrary and will not play an important role. Choosing a basis

$$|1\rangle \equiv |\uparrow\rangle, \quad |2\rangle \equiv |\downarrow\rangle, \quad (4.39)$$

the Hamiltonian, in this basis takes the form of the two-by-two matrix

$$H = \begin{pmatrix} E_0 & 0 \\ 0 & E_0 \end{pmatrix}. \quad (4.40)$$

The ability to tunnel must correspond to off-diagonal elements in the Hamiltonian matrix – there is no other option, in fact! So we must have a nonzero $H_{12} = \langle 1|H|2\rangle \neq 0$. Since the Hamiltonian must be hermitian, we must have $H_{12} = H_{21}^*$. For the time being we will take the off-diagonal elements to be real and therefore:

$$H_{12} = H_{21} = -\Delta, \quad \Delta > 0. \quad (4.41)$$

The sign of the real constant Δ is conventional. We could change it by a change of basis in which we let, for example $|2\rangle \rightarrow -|2\rangle$. Our choice will be convenient. The full Hamiltonian is now

$$H = \begin{pmatrix} E_0 & -\Delta \\ -\Delta & E_0 \end{pmatrix} = E_0 \mathbf{1} - \Delta \sigma_1, \quad (4.42)$$

where in the last step we wrote the matrix as a sum of a real number times the two-by-two identity matrix plus another real number times the first Pauli matrix. Both the identity matrix and the Pauli matrix are hermitian, consistent with having a hermitian Hamiltonian. The eigenvalues of the Hamiltonian follow from the equation

$$\begin{vmatrix} E_0 - \lambda & -\Delta \\ -\Delta & E_0 - \lambda \end{vmatrix} = 0 \quad \rightarrow \quad (E_0 - \lambda)^2 = \Delta^2 \quad \rightarrow \quad \lambda_{\pm} = E_0 \pm \Delta. \quad (4.43)$$

The eigenstates corresponding to these eigenvalues are

$$\begin{aligned} |G\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle), \quad E = E_0 - \Delta, \quad \text{Ground state} \\ |E\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle), \quad E = E_0 + \Delta, \quad \text{Excited state} \end{aligned} \quad (4.44)$$

Here $|G\rangle$ is the ground state (thus the G) and $|E\rangle$ is the excited state (thus the E). Our sign choices make the correspondence of the states with the wavefunction in figure 3 clear. The ground state $|G\rangle$ is a superposition, with the same sign, of the two localized (positive) wavefunctions. The excited state $|E\rangle$ has the wavefunction localized at z_0 (corresponding to $|1\rangle$) with a plus sign and the wavefunction localized at $-z_0$ (corresponding to $|2\rangle$) with a negative sign. Note that in the notation of section 3, the Hamiltonian in (4.42) corresponds to $\mathbf{g} = -\Delta \mathbf{x}$ and therefore $\mathbf{n} = -\mathbf{x}$ and $g = \Delta$. The excited state corresponds to the spin state $\frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)$, which points in the $-\mathbf{x}$ direction. The ground state corresponds to the spin state $\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$ which points in the $+\mathbf{x}$ direction.

The energy difference between these two eigenstates is 2Δ , which for the ammonia molecule takes the value

$$2\Delta = 0.9872 \times 10^{-4} \text{ ev}. \quad (4.45)$$

Setting this energy equal to the energy of a photon that may be emitted in such a transition we find

$$2\Delta = \hbar\omega = h\nu, \quad \nu = 23.870 \times 10^9 \text{ Hz} = 23.870 \text{ GHz}, \quad (4.46)$$

corresponding to a wavelength λ of

$$\lambda = 1.2559 \text{ cm}. \quad (4.47)$$

Let us consider the time evolution of an ammonia molecule that at $t = 0$ is in the state $|\uparrow\rangle$. Using (4.44) we express the initial state in terms of energy eigenstates:

$$|\Psi, 0\rangle = |\uparrow\rangle = \frac{1}{\sqrt{2}}(|G\rangle + |E\rangle). \quad (4.48)$$

The time-evolved state can now be readily written:

$$|\Psi, t\rangle = \frac{1}{\sqrt{2}}\left(e^{-i(E-\Delta)t/\hbar}|G\rangle + e^{-i(E+\Delta)t/\hbar}|E\rangle\right), \quad (4.49)$$

and we now rewrite it in terms of the $|\uparrow\rangle, |\downarrow\rangle$ states as follows

$$\begin{aligned} |\Psi, t\rangle &= \frac{1}{\sqrt{2}}\left(e^{-i(E-\Delta)t/\hbar}\frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) + e^{-i(E+\Delta)t/\hbar}\frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle)\right), \\ &= \frac{1}{2}e^{-iEt/\hbar}\left(e^{i\Delta t/\hbar}(|\uparrow\rangle + |\downarrow\rangle) + e^{-i\Delta t/\hbar}(|\uparrow\rangle - |\downarrow\rangle)\right), \\ &= e^{-iEt/\hbar}\left[\cos\left(\frac{\Delta t}{\hbar}\right)|\uparrow\rangle + i\sin\left(\frac{\Delta t}{\hbar}\right)|\downarrow\rangle\right]. \end{aligned} \quad (4.50)$$

The above time-dependent state oscillates from $|\uparrow\rangle$ to $|\downarrow\rangle$ with angular frequency $\omega = \Delta/\hbar \simeq 23 \text{ GHz}$. The probabilities $P_{\uparrow\downarrow}(t)$ that the state is found with nitrogen up or down are simply

$$\begin{aligned} P_{\uparrow}(t) &= |\langle \uparrow | \Psi, t \rangle|^2 = \cos^2\left(\frac{\Delta t}{\hbar}\right), \\ P_{\downarrow}(t) &= |\langle \downarrow | \Psi, t \rangle|^2 = \sin^2\left(\frac{\Delta t}{\hbar}\right). \end{aligned} \quad (4.51)$$

A plot of these is shown in Figure 4

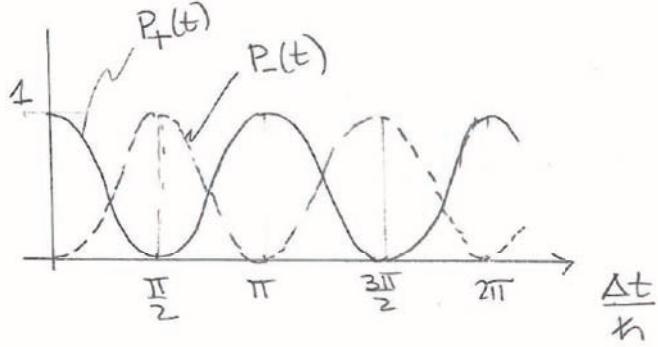


Figure 4: If the nitrogen atom starts in the state $|\uparrow\rangle$ at $t = 0$ (up position) then the probability $P_{\uparrow}(t)$ (shown as a continuous line) oscillates between one and zero. The probability $P_{\downarrow}(t)$ to be found in the state $|\downarrow\rangle$ is shown in dashed lines. Of course, $P_{\uparrow}(t) + P_{\downarrow}(t) = 1$. (Figure needs updating to change \pm to $\uparrow\downarrow$.)

5 Ammonia molecule in an electric field

Let us now consider the electrostatic properties of the ammonia molecule. The electrons tend to cluster towards the nitrogen, leaving the nitrogen vertex slightly negative and the hydrogen plane slightly positive. As a result we get an electric dipole moment μ that points down – when the nitrogen is up. The energy E of a dipole in an electric field \mathcal{E} is

$$E = -\mu \cdot \mathcal{E}. \quad (5.1)$$

With the electric field $\mathcal{E} = \mathcal{E}\mathbf{z}$ along the z axis, and $\mu = -\mu\mathbf{z}$, with $\mu > 0$, the state $|\uparrow\rangle$ with nitrogen up gets an extra positive contribution to the energy equal to $\mu\mathcal{E}$, while the $|\downarrow\rangle$ state gets the extra piece $-\mu\mathcal{E}$. The new Hamiltonian, including the effects of the electric field is then

$$H = \begin{pmatrix} E_0 + \mu\mathcal{E} & -\Delta \\ -\Delta & E_0 - \mu\mathcal{E} \end{pmatrix} = E_0 \mathbf{1} - \Delta \sigma_1 + \mu\mathcal{E} \sigma_3, \quad (5.2)$$

This corresponds to $g = \sqrt{(\mu\mathcal{E})^2 + \Delta^2}$ and therefore the energy eigenvalues are

$$\begin{aligned} E_E(\mathcal{E}) &= E_0 + \sqrt{\mu^2\mathcal{E}^2 + \Delta^2}, \\ E_G(\mathcal{E}) &= E_0 - \sqrt{\mu^2\mathcal{E}^2 + \Delta^2}, \end{aligned} \quad (5.3)$$

where we added the subscripts E for excited and G for ground to identify the energies as those of the excited and ground states when $\mathcal{E} = 0$. For small \mathcal{E} , or more precisely, small $\mu\mathcal{E}/\Delta$, we

have

$$\begin{aligned} E_E(\mathcal{E}) &\simeq E_0 + \Delta + \frac{\mu^2 \mathcal{E}^2}{2\Delta} + \mathcal{O}(\mathcal{E}^4), \\ E_G(\mathcal{E}) &\simeq E_0 - \Delta - \frac{\mu^2 \mathcal{E}^2}{2\Delta} + \mathcal{O}(\mathcal{E}^4), \end{aligned} \quad (5.4)$$

while for large $\mu\mathcal{E}$

$$\begin{aligned} E_E(\mathcal{E}) &\simeq E_0 + \mu\mathcal{E} + \mathcal{O}(1/\mathcal{E}), \\ E_G(\mathcal{E}) &\simeq E_0 - \mu\mathcal{E} + \mathcal{O}(1/\mathcal{E}). \end{aligned} \quad (5.5)$$

A plot of the energies is shown in Figure 5.

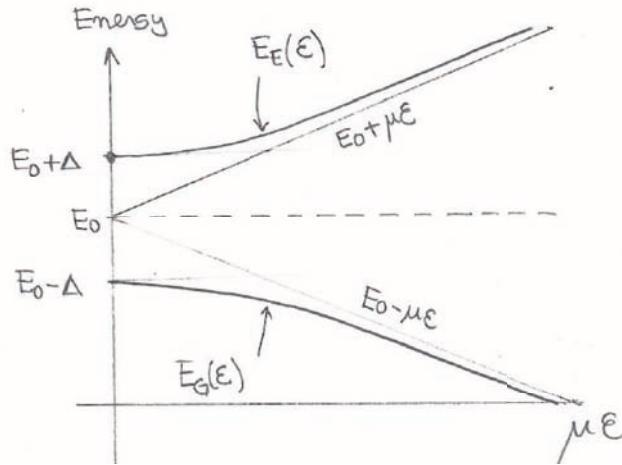


Figure 5: The energy levels of the two states of the ammonia molecule as a function of the magnitude \mathcal{E} of electric field. The higher energy state, with energy $E_E(\mathcal{E})$ coincides with $|E\rangle$ when $\mathcal{E} = 0$. The lower energy state, with energy $E_G(\mathcal{E})$ coincides with $|G\rangle$ when $\mathcal{E} = 0$.

This analysis gives us a way to split a beam of ammonia molecules into two beams, one with molecules in the state $|G\rangle$ and one with molecules in the state $|E\rangle$. As shown in the figure we have a beam entering a region with a spatially dependent electric field. The electric field gradient points up: the magnitude of the field is larger above than below. In a practical device $\mu\mathcal{E} \ll \Delta$ and we can use (5.4). A molecule in the $|E\rangle$ state will tend to go to the region of lower $|\mathcal{E}|$ as this is the region of low energy. Similarly a molecule in the $|G\rangle$ state will tend to go to the region of larger $|\mathcal{E}|$. Thus this device acts as a beam splitter.

The idea now is build a resonant electromagnetic cavity tuned to the frequency of 23.87 GHz and with very small losses (a high Q cavity). On one end, through a small hole, we let in a beam of ammonia molecules in the $|E\rangle$ state. These molecules exit the cavity through another

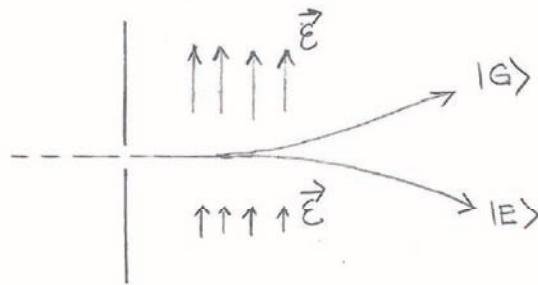


Figure 6: If a beam of ammonia molecules is exposed to an electric field with a strong gradient, molecules in the ground state $|G\rangle$ are deflected towards the stronger field (up) while molecules in the excited state $|E\rangle$ are deflected towards the weaker field (down).

hole on the opposite side (see Figure 7). If the design is done right, they exit on the ground state $|G\rangle$ thus having yielded an energy $2\Delta = \hbar\omega_0$ to the cavity. The ammonia molecules in the cavity interact with a spontaneously created electric field \mathcal{E} that oscillates with the resonant frequency. The interaction with such field induces the transition $|E\rangle \rightarrow |G\rangle$. This transition also feeds energy into the field. We want to understand this transition.

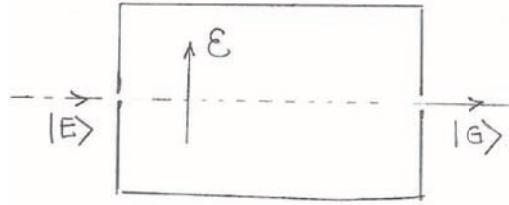


Figure 7: A resonant cavity tuned for 23.87GHz. A beam of ammonia molecules in the excited state $|E\rangle$ enter from the left. If properly designed, the molecules exit the cavity from the right on the ground state $|G\rangle$. In this process each molecule adds energy 2Δ to the electromagnetic field in the cavity.

The mathematics of the transition is clearer if we express the Hamiltonian in the primed basis

$$|1'\rangle \equiv |E\rangle, \quad |2'\rangle \equiv |G\rangle. \quad (5.6)$$

instead of the basis $|1\rangle = |\uparrow\rangle, |2\rangle = |\downarrow\rangle$ used to describe the Hamiltonian in (5.2). We can use

(4.44) to calculate the new matrix elements. For example

$$\begin{aligned}
\langle E|H|E\rangle &= \frac{1}{2}(\langle \uparrow | - \langle \downarrow |)H(|\uparrow\rangle - |\downarrow\rangle) \\
&= \frac{1}{2}(\langle \uparrow |H|\uparrow\rangle - \langle \uparrow |H|\downarrow\rangle - \langle \downarrow |H|\uparrow\rangle + \langle \downarrow |H|\downarrow\rangle) \\
&= \frac{1}{2}(E_0 + \mu\mathcal{E} - (-\Delta) - (-\Delta) + E_0 - \mu\mathcal{E}) \\
&= E_0 + \Delta,
\end{aligned} \tag{5.7}$$

and

$$\begin{aligned}
\langle E|H|G\rangle &= \frac{1}{2}(\langle \uparrow | - \langle \downarrow |)H(|+\rangle + |\downarrow\rangle) \\
&= \frac{1}{2}(\langle +|H|\uparrow\rangle + \langle \uparrow |H|\downarrow\rangle - \langle \downarrow |H|\uparrow\rangle - \langle \downarrow |H|\downarrow\rangle) \\
&= \frac{1}{2}(E_0 + \mu\mathcal{E} + (-\Delta) - (-\Delta) - (E_0 - \mu\mathcal{E})) \\
&= \mu\mathcal{E}.
\end{aligned} \tag{5.8}$$

and similarly $\langle G|H|G\rangle = E_0 - \Delta$. All in all the Hamiltonian in the new basis is given by

$$H = \begin{pmatrix} E_0 + \Delta & \mu\mathcal{E} \\ \mu\mathcal{E} & E_0 - \Delta \end{pmatrix} \quad \text{In the } |1'\rangle = |E\rangle, |2'\rangle = |G\rangle \text{ basis.} \tag{5.9}$$

We then write the wavefunction in terms of the amplitudes C_E and C_G to be in the $|E\rangle$ or $|G\rangle$ states respectively,

$$|\Psi\rangle = \begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix}. \tag{5.10}$$

The constant energy E_0 is not relevant – it can be set to any value and we choose the value zero. Doing so, the Schrödinger equation takes the form

$$i\hbar \frac{d}{dt} \begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix} = \begin{pmatrix} \Delta & \mu\mathcal{E} \\ \mu\mathcal{E} & -\Delta \end{pmatrix} \begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix}. \tag{5.11}$$

A strategy to solve this equation is to imagine that $\mu\mathcal{E}$ is very small compared to Δ so that

$$\begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix} = \begin{pmatrix} e^{-i\Delta t/\hbar} \beta_E(t) \\ e^{+i\Delta t/\hbar} \beta_G(t) \end{pmatrix}, \tag{5.12}$$

would be an exact solution with time-independent β_E and β_G if $\mu\mathcal{E} = 0$. When $\mu\mathcal{E}$ is small, we can expect solutions with β_E and β_G slowly varying in time (compared to the frequency Δ/\hbar) of the phases we have brought out to the open. We now substitute into (5.11), with the result (do the algebra!) with several terms canceling and

$$i\hbar \frac{d}{dt} \begin{pmatrix} \beta_E(t) \\ \beta_G(t) \end{pmatrix} = \begin{pmatrix} 0 & e^{i\omega_0 t} \mu\mathcal{E} \\ e^{-i\omega_0 t} \mu\mathcal{E} & 0 \end{pmatrix} \begin{pmatrix} \beta_E(t) \\ \beta_G(t) \end{pmatrix}, \quad \omega_0 \equiv \frac{2\Delta}{\hbar} \tag{5.13}$$

where we defined ω_0 as the frequency of a photon associated to the transition $|E\rangle \rightarrow |G\rangle$. This frequency is the resonant frequency of the cavity to be used. We now assume that in the cavity the electric field \mathcal{E} is at resonance so that

$$\mathcal{E}(t) = 2\mathcal{E}_0 \cos \omega_0 t = \mathcal{E}_0(e^{i\omega_0 t} + e^{-i\omega_0 t}) \quad (5.14)$$

so that

$$\begin{aligned} e^{i\omega_0 t} \mu \mathcal{E} &= \mu \mathcal{E}_0 (1 + e^{2i\omega_0 t}) \\ e^{-i\omega_0 t} \mu \mathcal{E} &= \mu \mathcal{E}_0 (1 + e^{-2i\omega_0 t}) \end{aligned} \quad (5.15)$$

We can now go back to the differential equation which gives

$$\begin{aligned} i \dot{\beta}_E(t) &= \frac{\mu \mathcal{E}_0}{\hbar} (1 + e^{2i\omega_0 t}) \beta_G(t) \\ i \dot{\beta}_G(t) &= \frac{\mu \mathcal{E}_0}{\hbar} (1 + e^{-2i\omega_0 t}) \beta_E(t). \end{aligned} \quad (5.16)$$

With $\mu \mathcal{E}_0$ small, the rates of change of β_E and β_G will necessarily be small, as $\mu \mathcal{E}_0$ appears multiplicatively on the right-hand side. Thus β_E and β_G are essentially constant during the period of oscillation of the exponential terms $e^{\pm 2i\omega_0 t}$. Since these exponentials have zero time-averaged values, they can be dropped. Thus, we get

$$\begin{aligned} \dot{\beta}_E(t) &= -i \frac{\mu \mathcal{E}_0}{\hbar} \beta_G(t) \\ \dot{\beta}_G(t) &= -i \frac{\mu \mathcal{E}_0}{\hbar} \beta_E(t). \end{aligned} \quad (5.17)$$

Taking another time derivative of the top equation we find;

$$\ddot{\beta}_E(t) = -\left(\frac{\mu \mathcal{E}_0}{\hbar}\right)^2 \beta_E(t) \quad (5.18)$$

This has the simple solution in terms of sines and cosines. If we assume that the molecule at time equal zero is indeed in the state $|E\rangle$ we then write

$$\beta_E(t) = \cos\left(\frac{\mu \mathcal{E}_0}{\hbar} t\right) \rightarrow \beta_G(t) = -i \sin\left(\frac{\mu \mathcal{E}_0}{\hbar} t\right). \quad (5.19)$$

The time dependent probability $P_E(t)$ to be in the $|E\rangle$ state is then

$$P_E(t) = |C_E(t)|^2 = \left|e^{-i\Delta t/\hbar} \beta_E(t)\right|^2 = \cos^2\left(\frac{\mu \mathcal{E}_0}{\hbar} t\right). \quad (5.20)$$

This is our desired result. The molecule that enters the cavity in the state $|E\rangle$ will leave the cavity in the state $|G\rangle$ if the travel time T is such that the probability $P_E(T)$ to be in $|E\rangle$ vanishes. For this we need

$$\cos\left(\frac{\mu \mathcal{E}_0}{\hbar} T\right) = 0 \rightarrow \frac{\mu \mathcal{E}_0}{\hbar} T = \frac{\pi}{2}, \frac{3\pi}{2}, \dots \quad (5.21)$$

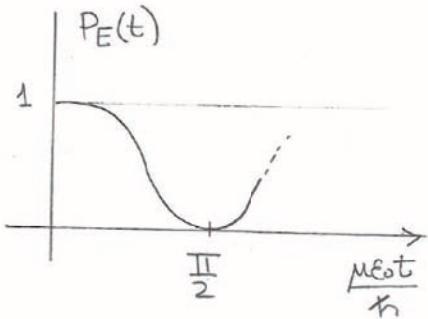


Figure 8: An ammonia molecule enters the resonant cavity at $t = 0$ in the excited state $|E\rangle$. We show probability $P_E(t)$ for the ammonia molecule to be in the excited $|E\rangle$ at time t . If the molecule exits at a time T for which $P_E(T) = 0$, the molecule will be in the ground state $|G\rangle$, as desired.

See figure 8.

If the velocity of the ammonia molecules is adjusted for this to happen, each molecule gives energy 2Δ to the cavity's electromagnetic field. The cavity's EM field, by producing the transition $|E\rangle \rightarrow |G\rangle$ of the traveling molecules, stimulates the emission of radiation. Moreover the energy released is in a field with the same configuration and frequency as the stimulating EM field. The molecules thus help build a *coherent* field in the cavity. Such a cavity is then a MASER, an acronym that stands for Microwave Amplification by Stimulated Emission of Radiation. The molecules are uncharged and therefore their motions produce no unwanted EM signals – this is in contrast to electrons in vacuum amplifiers, which produce *shot* noise.

Charles H. Townes, James P. Gordon, and H. J. Zeiger built the first ammonia maser working at Columbia University in 1953. As stated by Charles H. Townes in his Nobel lecture on December 11, 1964, “masers yield the most perfect amplification allowed by the uncertainty principle”. For such EM waves the uncertainty principle can be written as

$$\Delta n \Delta \phi \geq \frac{1}{2},$$

where Δn is the uncertainty in the number of photons in the field and $\Delta \phi$ is the phase uncertainty of the wave, in radians. For a coherent field $\Delta n = \sqrt{\bar{n}}$, with \bar{n} the expected number of photons. The saturation of the uncertainty principle leads to a phase uncertainty

$$\Delta \phi = \frac{1}{2\sqrt{\bar{n}}}, \quad (5.22)$$

that for any realistic \bar{n} is fantastically small.

6 Nuclear Magnetic Resonance

The problem we want to discuss is that of a spin in a time dependent magnetic field. This magnetic field has a time-independent z component and a circularly polarized field representing a magnetic field rotating on the (x, y) plane. More concretely, we have

$$\mathbf{B}(t) = B_0 \mathbf{z} + B_1 (\mathbf{x} \cos \omega t - \mathbf{y} \sin \omega t). \quad (6.23)$$

Typically, the constant z -component B_0 is larger than B_1 , the magnitude of the RF (radio-frequency) signal. The time dependent part of the field points along the x axis at $t = 0$ and is rotating with angular velocity ω in the clockwise direction of the (x, y) plane. The spin Hamiltonian is

$$H_S(t) = -\gamma \mathbf{B}(t) \cdot \mathbf{S} = -\gamma \left(B_0 \hat{S}_z + B_1 (\cos \omega t \hat{S}_x - \sin \omega t \hat{S}_y) \right). \quad (6.24)$$

Not only is this Hamiltonian time dependent, but the Hamiltonian at different times do not commute. So this is a nontrivial time evolution problem!

We attempt to simplify the problem by considering a frame of reference that rotates with the magnetic field. For this, imagine first the case when $H_S = 0$ because the magnetic field is zero. With no magnetic field spin states would simply be static. What would the Hamiltonian be in the frame rotating about the z -axis with angular frequency ω , just like the magnetic field above? In that frame, the spin states that are fixed in the original frame would be seen to rotate with positive angular velocity ω about the z direction. There must be a Hamiltonian that does have that effect. Since the unitary operator \mathcal{U} that generates this rotation is

$$\mathcal{U}(t) = \exp\left(-\frac{i\omega t \hat{S}_z}{\hbar}\right) \rightarrow H_{\mathcal{U}} = \omega \hat{S}_z, \quad (6.25)$$

where the expression $H_{\mathcal{U}}$ for the Hamiltonian in the rotating frame is read from the relation $\mathcal{U} = \exp(-iH_{\mathcal{U}}t/\hbar)$. For the original case, when the original Hamiltonian in the static frame is H_S , we will use the above operator \mathcal{U} to define a new rotating-frame state $|\Psi_R\rangle$ as follows

$$|\Psi_R, t\rangle \equiv \mathcal{U}(t)|\Psi, t\rangle. \quad (6.26)$$

If we knew $|\Psi_R, t\rangle$ we would know $|\Psi, t\rangle$. We wish to find out if the Schrödinger equation for $|\Psi_R\rangle$ becomes simpler. For this we must determine the corresponding Hamiltonian H_R . One quick way to do this is to note that the above equation implies that

$$|\Psi_R, t\rangle \equiv \mathcal{U}(t)\mathcal{U}_S(t)|\Psi, 0\rangle. \quad (6.27)$$

where we have introduced the unitary operator $\mathcal{U}_S(t)$ associated with the Hamiltonian $H_S(t)$ that evolves $|\Psi\rangle$ in time. Since the Hamiltonian associated to an arbitrary unitary time-evolution operator \mathcal{U} is $i\hbar(\partial_t \mathcal{U})\mathcal{U}^\dagger$ (if you don't recall this, derive it from a Schrödinger equation)

we have

$$\begin{aligned}
H_R &= i\hbar \partial_t (\mathcal{U} \mathcal{U}_S) \mathcal{U}_S^\dagger \mathcal{U}^\dagger \\
&= i\hbar (\partial_t \mathcal{U}) \mathcal{U}^\dagger + \mathcal{U} i\hbar (\partial_t \mathcal{U}_S) \mathcal{U}_S^\dagger \mathcal{U}^\dagger \\
\rightarrow H_R &= H_{\mathcal{U}} + \mathcal{U} H_S \mathcal{U}^\dagger,
\end{aligned} \tag{6.28}$$

where $H_{\mathcal{U}}$ is the Hamiltonian associated to \mathcal{U} . Since \mathcal{U} is the one above, we have $H_{\mathcal{U}} = \omega \hat{S}_z$ and therefore,

$$\begin{aligned}
H_R &= \omega \hat{S}_z + \exp\left(-\frac{i\omega t \hat{S}_z}{\hbar}\right) \left[-\gamma \left(B_0 \hat{S}_z + B_1 (\cos \omega t \hat{S}_x - \sin \omega t \hat{S}_y) \right) \right] \exp\left(\frac{i\omega t \hat{S}_z}{\hbar}\right) \\
&= (-\gamma B_0 + \omega) \hat{S}_z - \gamma B_1 \underbrace{\exp\left(-\frac{i\omega t \hat{S}_z}{\hbar}\right) (\cos \omega t \hat{S}_x - \sin \omega t \hat{S}_y) \exp\left(\frac{i\omega t \hat{S}_z}{\hbar}\right)}_{\hat{M}(t)}
\end{aligned} \tag{6.29}$$

and the big question is what is $\hat{M}(t)$. We can proceed by calculating the time-derivative of \hat{M} :

$$\begin{aligned}
\partial_t \hat{M} &= e^{-\frac{i\omega t \hat{S}_z}{\hbar}} \left(-\frac{i\omega}{\hbar} [\hat{S}_z, \cos \omega t \hat{S}_x - \sin \omega t \hat{S}_y] + [-\omega \sin \omega t \hat{S}_x - \omega \cos \omega t \hat{S}_y] \right) e^{\frac{i\omega t \hat{S}_z}{\hbar}} \\
&= e^{-\frac{i\omega t \hat{S}_z}{\hbar}} \left(-\frac{i\omega}{\hbar} [i\hbar \cos \omega t \hat{S}_y + i\hbar \sin \omega t \hat{S}_x] + [-\omega \sin \omega t \hat{S}_x - \omega \cos \omega t \hat{S}_y] \right) e^{\frac{i\omega t \hat{S}_z}{\hbar}} \\
&= e^{-\frac{i\omega t \hat{S}_z}{\hbar}} \left(\omega \cos \omega t \hat{S}_y + \omega \sin \omega t \hat{S}_x - \omega \sin \omega t \hat{S}_x - \omega \cos \omega t \hat{S}_y \right) e^{\frac{i\omega t \hat{S}_z}{\hbar}} \\
&= 0.
\end{aligned} \tag{6.30}$$

This is very good news. Since \hat{M} has no time dependence, we can evaluate it at any time. The simplest time is $t = 0$ and we thus find that

$$\hat{M}(t) = \hat{S}_x. \tag{6.31}$$

As a result we have a totally time-independent H_R :

$$\begin{aligned}
H_R &= (-\gamma B_0 + \omega) \hat{S}_z - \gamma B_1 \hat{S}_x \\
&= -\gamma \left[\left(B_0 - \frac{\omega}{\gamma} \right) \hat{S}_z + B_1 \hat{S}_x \right] \\
&= -\gamma \left[B_0 \left(1 - \frac{\omega}{\omega_0} \right) \hat{S}_z + B_1 \hat{S}_x \right],
\end{aligned} \tag{6.32}$$

using $\omega_0 = \gamma B_0$ for the Larmor frequency associated with the constant component of the field. We thus have a Hamiltonian H_R that can be associated with a magnetic field \mathbf{B}_R given by Setting

$$H_R = -\gamma \mathbf{B}_R \cdot \mathbf{S}, \quad \rightarrow \quad \mathbf{B}_R = B_1 \mathbf{x} + B_0 \left(1 - \frac{\omega}{\omega_0} \right) \mathbf{z}. \tag{6.33}$$

The full solution for the state is obtained beginning with (6.26) and (6.25):

$$|\Psi, t\rangle = \mathcal{U}^\dagger(t)|\Psi_R, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right]|\Psi_R, t\rangle. \quad (6.34)$$

Since H_R is a time-independent Hamiltonian, we have that the full time evolution is

$$|\Psi, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right] \exp\left[-i\frac{(-\gamma \mathbf{B}_R \cdot \mathbf{S})t}{\hbar}\right] |\Psi, 0\rangle \quad (6.35)$$

where the solution for $|\Psi_R, t\rangle$ is simply the time evolution from the $t = 0$ state generated by the Hamiltonian H_R . We have thus found that

$$|\Psi, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right] \exp\left[i\frac{\gamma \mathbf{B}_R \cdot \mathbf{S} t}{\hbar}\right] |\Psi, 0\rangle. \quad (6.36)$$

Exercise. Verify that for $B_1 = 0$ the above solution reduces to the one describing precession about the z -axis.

In the applications to be discussed below we always have

$$B_1 \ll B_0. \quad (6.37)$$

Consider now the evolution of a spin that initially points in the positive z direction. We look at two cases:

- $\omega \ll \omega_0$. In this case from (6.33) we have

$$\mathbf{B}_R \simeq B_0 \mathbf{z} + B_1 \mathbf{x}. \quad (6.38)$$

This is a field mostly along the z axis, but tipped a little towards the x axis. The right-most exponential in (6.36) makes the spin precess about the direction \mathbf{B}_R quite quickly, for $|\mathbf{B}_R| \sim B_0$ and thus the angular rate of precession is pretty much ω_0 . The next exponential in (6.36) induces a rotation about the z -axis with smaller angular velocity ω . This is shown in the figure.

- $\omega = \omega_0$. In this case from (6.33) we have

$$\mathbf{B}_R = B_1 \mathbf{x}. \quad (6.39)$$

In this case the right-most exponential in (6.36) makes the spin precess about the x axis it goes from the z axis towards the y axis with angular velocity $\omega_1 = \gamma B_1$. If we time the RF signal to last a time T such that

$$\omega_1 T = \frac{\pi}{2} \quad \rightarrow \quad T = \frac{\pi}{2\gamma B_1}, \quad (6.40)$$

the state $|\Psi_R, T\rangle$ points along the y axis. The effect of the other exponential in (6.36) is just to rotate the spin on the (x, y) plane. We have

$$|\Psi, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right] |\Psi_R, t\rangle, \quad t < T, \quad (6.41)$$

and if the RF pulse turns off after time T ,

$$|\Psi, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right] |\Psi_R, T\rangle, \quad t > T, \quad (6.42)$$

The state $|\Psi, t\rangle$ can be visualized as a spin that is slowly rotating with angular velocity ω_1 from the z axis towards the plane, while rapidly rotating around the z axis with angular velocity ω . As a result the tip of the spin is performing a spiral motion on the surface of a hemisphere. By the time the polar angle reaches $\pi/2$ the RF signal turns off and the spin now just rotates within the (x, y) plane. This is called a 90° pulse.

Magnetic Resonance Imaging Based on work on nuclear magnetic resonance by Edward Purcell, who worked at MIT's radiation laboratory, and Felix Bloch. They got the Nobel prize for this work in 1952.

The NMR work led to the development of the technique called MRI for Magnetic Resonance Imaging. First studies in humans was done in 1977. The new perspective, as compared to X-rays, was that MRI allows one to distinguish various soft tissues.

The human body is mostly composed of water molecules. In those we have many hydrogen atoms, whose nuclei are protons and are the main players through their magnetic dipole moments (nuclear dipoles).

With a given external and large constant magnetic field B_0 , at a given temperature, there is a net alignment of nuclear spins along B_0 . This is called the “longitudinal magnetization”. For all intents and purposes we have a net number of spins in play.

We apply a 90° pulse so that we get the spins to rotate with Larmor frequency ω_0 in the (x, y) plane. These rotating dipoles produce an oscillating magnetic field which is a signal that can be picked up by a receiver. The magnitude of the signal is proportional to the proton density. This is the first piece of information and allows differentiation of tissues.

The above signal from the rotation of the spins decays with a time constant T_2 that can be measured and is typically much smaller than a second. This decay is attributed to interactions between the spins. A T_2 weighted image allows doctors to detect abnormal accumulation of fluids (edema).

There is another time constant T_1 (of order one second) that controls the time to regain the longitudinal magnetization. This effect is due to the spins interacting with the rest of

the lattice of atoms. White matter, grey matter, and cerebrospinal fluids are distinguished by different T_1 constants (they have about the same proton density).

MRI's commonly include the use of contrast agents, which are substances that shorten the time constant T_1 and are usually administered by injection into the blood stream. The contrast agent (gadolinium) can accumulate at organs or locations where information is valuable. For a number of substances one can use the MRI apparatus to determine their (T_1, T_2) constants and build a table of data. This table can then be used as an aid to identify the results of MRI's.

The typical MRI machine has a B_0 of about 2T (two tesla or 20,000 gauss). This requires a superconducting magnet with cooling by liquid helium. For people with claustrophobia there are “open” MRI scanners that work with lower magnetic fields. In addition there are about three *gradient* magnets, each of about 200 gauss. They change locally the value of B_0 and thus provide spatial resolution as the Larmor frequency becomes spatially dependent. One can thus attain spatial resolutions of about half a millimiter! MRI's are considered safe, as there is no evidence of biological harm caused by very large static magnetic fields.

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Chapter 1: Key Features of Quantum Mechanics

Quantum mechanics is now almost one-hundred years old, but we are still discovering some of its surprising features and it remains the subject of much investigation and speculation. The framework of quantum mechanics is a rich and elegant extension of the framework of classical physics. It is also counterintuitive and almost paradoxical.

Quantum physics has replaced classical physics as the correct fundamental description of our physical universe. It is used routinely to describe most phenomena that occur at short distances. Quantum physics is the result of applying the framework of quantum mechanics to different physical phenomena. We thus have Quantum Electrodynamics, when quantum mechanics is applied to electromagnetism, Quantum Optics, when it is applied to light and optical devices, or Quantum Gravity, when it is applied to gravitation. Quantum mechanics indeed provides a remarkably coherent and elegant framework. The era of quantum physics begins in 1925, with the discoveries of Schrödinger and Heisenberg. The seeds for these discoveries were planted by Planck, Einstein, Bohr, de Broglie, and others. It is a tribute to human imagination that we have been able to discover the counterintuitive and abstract set of rules that define quantum mechanics. Here we aim to explain and provide some perspective on the main features of this framework.

We will begin by discussing the property of linearity, which quantum mechanics shares with electromagnetic theory. This property tells us what kind of theory quantum mechanics is and why, it could be argued, it is simpler than classical mechanics. We then turn to photons, the particles of light. We use photons and polarizers to explain why quantum physics is not deterministic and, in contrast with classical physics, the results of some experiments cannot be predicted. Quantum mechanics is a framework in which we can only predict the *probabilities* for the various outcomes of any given experiment. Our next subject is quantum superpositions, in which a quantum object somehow manages to exist simultaneously in two mutually incompatible states. A quantum light-bulb, for example, could be in a state in which it is both on and off at the same time!

1 Linearity of the equations of motion

In physics a theory is usually described by a set of equations for some quantities called the **dynamical variables** of the theory. After writing a theory, the most important task is finding solutions of the equations. A solution of the equations describes a possible reality, according to the theory. Because an expanding universe is a solution of Albert Einstein's gravitational equations, for example, it follows that an expanding universe is possible, according to this theory. A single theory may have many solutions, each describing a possible reality.

There are linear theories and nonlinear theories. Nonlinear theories are more complex than linear theories. In a linear theory a remarkable fact takes place: if you have two solutions you obtain a third solution of the theory simply by adding the two solutions. An example of a beautiful linear theory is Maxwell's theory of electromagnetism, a theory that governs the behavior of electric and magnetic fields. A field, as you probably know, is a quantity whose values may depend on position and on time. A simple solution of this theory describes an electromagnetic wave propagating in a given direction. Another simple solution could describe an electromagnetic wave propagating in a

different direction. Because the theory is linear, having the two waves propagating simultaneously, each in its own direction and without affecting each other, is a new and consistent solution. The sum is a solution in the sense that the electric field in the new solution is the sum of the electric field in the first solution plus the electric field in the second solution. The same goes for the magnetic field: the magnetic field in the new solution is the sum of the magnetic field in the first solution plus the magnetic field in the second solution. In fact you can add any number of solutions to still find a solution. Even if this sounds esoteric, you are totally familiar with it. The air around you is full of electromagnetic waves, each one propagating oblivious to the other ones. There are the waves of thousands of cell phones, the waves carrying hundreds of wireless internet messages, the waves from a plethora of radio-stations, TV stations, and many, many more. Today, a single transatlantic cable can carry simultaneously millions of telephone calls, together with huge amounts video and internet data. All of that courtesy of linearity.

More concretely, we say that Maxwell's equations are **linear** equations. A solution of Maxwell's equation is described by an electric field \mathbf{E} , a magnetic field \mathbf{B} , a charge density ρ and a current density \mathbf{J} , all collectively denoted as $(\mathbf{E}, \mathbf{B}, \rho, \mathbf{J})$. This collection of fields and sources satisfy Maxwell's equations. Linearity implies that if $(\mathbf{E}, \mathbf{B}, \rho, \mathbf{J})$ is a solution so is $(\alpha\mathbf{E}, \alpha\mathbf{B}, \alpha\rho, \alpha\mathbf{J})$, where all fields and sources have been multiplied by the constant α . Given two solutions

$$(\mathbf{E}_1, \mathbf{B}_1, \rho_1, \mathbf{J}_1), \quad \text{and} \quad (\mathbf{E}_2, \mathbf{B}_2, \rho_2, \mathbf{J}_2), \quad (1.1)$$

linearity also implies that we can obtain a new solution by adding them

$$(\mathbf{E}_1 + \mathbf{E}_2, \mathbf{B}_1 + \mathbf{B}_2, \rho_1 + \rho_2, \mathbf{J}_1 + \mathbf{J}_2). \quad (1.2)$$

The new solution may be called the superposition of the two original solutions.

It is not hard to explain what is, in general, a linear equation or a linear set of equations. Consider the equation

$$L u = 0, \quad (1.3)$$

where, schematically, u denotes the unknown. The unknown may be a number, or a function of time, a function of space, a function of time and space, essentially anything unknown! In fact, u could represent a collection of unknowns, in which case we would replace u above by u_1, u_2, \dots . The symbol L denotes a **linear operator**, an object that satisfies the following two properties

$$L(u_1 + u_2) = Lu_1 + Lu_2, \quad L(a u) = aLu, \quad (1.4)$$

where a is a number. Note that these conditions imply that

$$L(\alpha u_1 + \beta u_2) = \alpha Lu_1 + \beta Lu_2, \quad (1.5)$$

showing that if u_1 is a solution ($Lu_1 = 0$) and u_2 is a solution ($Lu_2 = 0$) then $\alpha u_1 + \beta u_2$ is also a solution. We call $\alpha u_1 + \beta u_2$ the **general superposition** of the solutions u_1 and u_2 . An example may help. Consider the equation

$$\frac{du}{dt} + \frac{1}{\tau} u = 0, \quad (1.6)$$

where τ is a constant with units of time. This is, in fact, a linear differential equation, and takes the form $L u = 0$ if we define

$$L u \equiv \frac{du}{dt} + \frac{1}{\tau} u \quad (1.7)$$

Exercise 1. Verify that (1.7) satisfies the conditions for a linear operator.

Einstein's theory of general relativity is a nonlinear theory whose dynamical variable is a gravitational field, the field that describes, for example, how planets move around a star. Being a nonlinear theory, you simply cannot add the gravitational fields of different solutions to find a new solution. This makes Einstein's theory rather complicated, by all accounts much more complicated than Maxwell theory. In fact, classical mechanics, as invented mostly by Isaac Newton, is also a nonlinear theory! In classical mechanics the dynamical variables are positions and velocities of particles, acted by forces. There is no general way to use two solutions to build a third.

Indeed, consider the equation of motion for a particle on a line under the influence of a time-independent potential $V(x)$, which is in general an arbitrary function of x . The dynamical variable in this problem is $x(t)$, the position as a function of time. Letting V' denote the derivative of V with respect to its argument, Newton's second law takes the form

$$m \frac{d^2x(t)}{dt^2} = -V'(x(t)). \quad (1.8)$$

The left-hand side is the mass times acceleration and the right hand side is the force experienced by the particle in the potential. It is probably worth to emphasize that the right hand side is the function $V'(x)$ evaluated for x set equal to $x(t)$:

$$V'(x(t)) \equiv \left. \frac{\partial V(x)}{\partial x} \right|_{x=x(t)}. \quad (1.9)$$

While we could have used here an ordinary derivative, we wrote a partial derivative as is commonly done for the general case of time dependent potentials. The reason equation (1.8) is not a linear equation is that the function $V'(x)$ is not linear. In general, for arbitrary functions u and v we expect

$$V'(au) \neq aV'(u), \quad \text{and} \quad V'(u+v) \neq V'(u) + V(v). \quad (1.10)$$

As a result given a solution $x(t)$, the scaled solution $\alpha x(t)$ is not expected to be a solution. Given two solutions $x_1(t)$ and $x_2(t)$ then $x_1(t) + x_2(t)$ is not guaranteed to be a solution either.

Exercise. What is the most general potential $V(x)$ for which the equation of motion for $x(t)$ is linear?

Quantum mechanics is a linear theory. The signature equation in this theory, the so-called Schrödinger equation is a linear equation for a quantity called the **wavefunction** and it determines its time evolution. The wavefunction is the dynamical variable in quantum mechanics but, curiously, its physical interpretation was not clear to Erwin Schrödinger when he wrote the equation in 1925. It was Max Born, who months later suggested that the wavefunction encodes probabilities. This was the correct physical interpretation, but it was thoroughly disliked by many, including Schrödinger, who remained unhappy about it for the rest of his life. The linearity of quantum mechanics implies a profound simplicity. In some sense quantum mechanics is simpler than classical mechanics. In quantum mechanics solutions can be added to form new solutions.

The wavefunction Ψ depends on time and may also depend on space. The Schrödinger equation (SE) is a partial differential equation that takes the form

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi, \quad (1.11)$$

where the Hamiltonian (or energy operator) \hat{H} is a linear operator that can act on wavefunctions:

$$\hat{H}(a\Psi) = a\hat{H}\Psi, \quad \hat{H}(\Psi_1 + \Psi_2) = \hat{H}(\Psi_1) + \hat{H}(\Psi_2), \quad (1.12)$$

with a a constant that in fact need not be real; it can be a complex number. Of course, \hat{H} itself does not depend on the wavefunction! To check that the Schrödinger equation is linear we cast it in the form $L\Psi = 0$ with L defined as

$$L\Psi \equiv i\hbar \frac{\partial \Psi}{\partial t} - \hat{H}\Psi \quad (1.13)$$

It is now a simple matter to verify that L is a linear operator. Physically this means that if Ψ_1 and Ψ_2 are solutions to the Schrödinger equation, then so is the superposition $\alpha\Psi_1 + \beta\Psi_2$, where α and β are both complex numbers, i.e. ($\alpha, \beta \in \mathbb{C}$)

2 Complex Numbers are Essential

Quantum mechanics is the first physics theory that truly makes use of *complex* numbers. The numbers most of us use for daily life (integers, fractions, decimals) are *real* numbers. The set of complex numbers is denoted by \mathbb{C} and the set of real numbers is denoted by \mathbb{R} . Complex numbers appear when we combine real numbers with the imaginary unit i , defined to be equal to the square root of minus one: $i \equiv \sqrt{-1}$. Being the square root of minus one, it means that i squared must give minus one: $i^2 = -1$. Complex numbers are fundamental in mathematics. An equation like $x^2 = -4$, for an unknown x cannot be solved if x has to be real. No real number squared gives you minus one. But if we allow for complex numbers, we have the solutions $x = \pm 2i$. Mathematicians have shown that all polynomial equations can be solved in terms of complex numbers.

A complex number z , in all generality, is a number of the form

$$z = a + ib \in \mathbb{C}, \quad a, b \in \mathbb{R}. \quad (2.1)$$

Here a and b are real numbers, and ib denotes the product of i with b . The number a is called the real part of z and b is called the imaginary part of z :

$$\operatorname{Re} z = a, \quad \operatorname{Im} z = b. \quad (2.2)$$

The complex conjugate z^* of z is defined by

$$z^* = a - ib. \quad (2.3)$$

You can quickly verify that a complex number z is real if $z^* = z$ and it is purely imaginary if $z^* = -z$. For any complex number $z = a + ib$ one can define the *norm* $|z|$ of the complex number to be a *positive, real* number given by

$$|z| = \sqrt{a^2 + b^2}. \quad (2.4)$$

You can quickly check that

$$|z|^2 = zz^*, \quad (2.5)$$

where $z^* \equiv a - ib$ is called the complex conjugate of $z = a + ib$. Complex numbers are represented as vectors in a two dimensional “complex plane”. The real part of the complex number is the x component of the vector and the imaginary part of the complex number is the y component. If you consider the unit length vector in the complex plane making an angle θ with the x axis has x component $\cos \theta$ and

y component $\sin \theta$. The vector is therefore the complex number $\cos \theta + i \sin \theta$. Euler's identity relates this to the exponential of $i\theta$:

$$e^{i\theta} = \cos \theta + i \sin \theta. \quad (2.6)$$

A complex number of the form $e^{i\chi}$, with χ real is called a *pure phase*.

While complex numbers are sometimes useful in classical mechanics or Maxwell theory, they are not strictly needed. None of the dynamical variables, which correspond to measurable quantities, is a complex number. In fact, complex numbers can't be measured at all: all measurements in physics result in real numbers. In quantum mechanics, however, complex numbers are fundamental. The Schrödinger equation involves complex numbers. Even more, the wavefunction, the dynamical variable of quantum mechanics it itself a complex number:

$$\Psi \in \mathbb{C}. \quad (2.7)$$

Since complex numbers cannot be measured the relation between the wavefunction and a measurable quantity must be somewhat indirect. Born's idea to identify probabilities, which are always positive real numbers, with the square of the norm of the wavefunction was very natural. If we write the wavefunction of our quantum system as Ψ , the probabilities for possible events are computed from $|\Psi|^2$. The mathematical framework required to express the laws of quantum mechanics consists of complex vector spaces. In any vector space we have objects called vectors that can be added together. In a complex vector space a vector multiplied by a complex number is still a vector. As we will see in our study of quantum mechanics it is many times useful to think of the wavefunction Ψ as a vector in some complex vector space.

3 Loss of Determinism

Maxwell's crowning achievement was the realization that his equations of electromagnetism allowed for the existence of propagating waves. In particular, in 1865 he conjectured that light was an electromagnetic wave, a propagating fluctuation of electric and magnetic fields. He was proven right in subsequent experiments. Towards the end of the nineteenth century physicists were convinced that light was a wave. The certainty, however, did not last too long. Experiments on blackbody radiation and on the photo-emission of electrons suggested that the behavior of light had to be more complicated than that of a simple wave. Max Planck and Albert Einstein were the most prominent contributors to the resolution of the puzzles raised by those experiments.

In order to explain the features of the photoelectric effect, Einstein postulated (1905) that in a light beam the energy comes in quanta – the beam is composed of packets of energy. Einstein essentially implied that light was made up of particles, each carrying a fixed amount of energy. He himself found this idea disturbing, convinced like most other contemporaries that, as Maxwell had shown, light was a wave. He anticipated that a physical entity, like light, that could behave both as a particle and as a wave could bring about the demise of classical physics and would require a completely new physical theory. He was in fact right. Though he never quite liked quantum mechanics, his ideas about particles of light, later given the name *photons*, helped construct this theory.

It took physicists until 1925 to accept that light could behave like a particle. The experiments of Arthur Compton (1923) eventually convinced most skeptics. Nowadays, particles of light, or photons, are routinely manipulated in laboratories around the world. Even if mysterious, we have grown accustomed to them. Each photon of visible light carries very little energy – a small laser pulse can

contain many billions of photons. Our eye, however, is a very good photon detector: in total darkness, we are able to see light when as little as ten photons hit upon our retina. When we say that light behaves like a particle we mean a quantum mechanical particle: a packet of energy and momentum that is not composed of smaller packets. We *do not* mean a classical point particle or Newtonian corpuscle, which is a zero-size object with definite position and velocity.

As it turns out, the energy of a photon depends only on the color of the light. As Einstein discovered the energy E and frequency ν for a photon are related by

$$E = h\nu \quad (3.1)$$

The frequency of a photon determines the wavelength λ of the light through the relation $\nu\lambda = c$, where c is the speed of light. All green photons, for example, have the same energy. To increase the energy in a light beam while keeping the same color, one simply needs more photons.

As we now explain, the existence of photons implies that Quantum Mechanics is not deterministic. By this we mean that the result of an experiment cannot be determined, as it would in classical physics, by the conditions that are under the control of the experimenter.

Consider a polarizer whose preferential direction is aligned along the \hat{x} direction, as shown in Figure 1. Light that is linearly polarized along the \hat{x} direction namely, light whose electric field points in this direction, goes through the polarizer. If the incident light polarization is orthogonal to the \hat{x} direction the light will not go through at all. Thus light linearly polarized in the \hat{y} direction will be totally absorbed by the polarizer. Now consider light polarized along a direction forming an angle α with the x -axis, as shown in Figure 2. What happens?

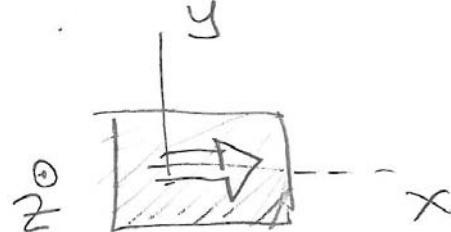


Figure 1: A polarizer that transmits light linearly polarized along the \hat{x} direction.

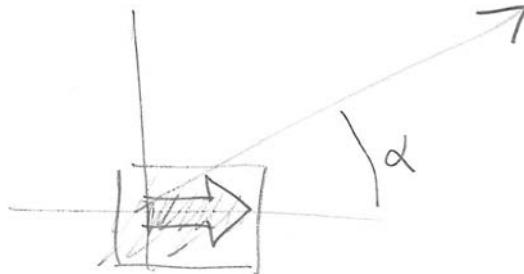


Figure 2: Light linearly polarized along the direction at an angle α hitting the polarizer.

Thinking of the light as a propagating wave, the incident electric field \mathbf{E}_α makes an angle α with the x -axis and therefore takes the form

$$\mathbf{E}_\alpha = E_0 \cos \alpha \hat{x} + E_0 \sin \alpha \hat{y}. \quad (3.2)$$

This is an electric field of magnitude E_0 . In here we are ignoring the time and space dependence of the wave; they are not relevant to our discussion. When this electric field hits the polarizer, the component along $\hat{\mathbf{x}}$ goes through and the component along $\hat{\mathbf{y}}$ is absorbed. Thus

$$\text{Beyond the polarizer: } \mathbf{E} = E_0 \cos \alpha \hat{\mathbf{x}}. \quad (3.3)$$

You probably recall that the energy in an electromagnetic wave is proportional to the square of the magnitude of the electric field. This means that the fraction of the beam's energy that goes through the polarizer is $(\cos \alpha)^2$. It is also well known that the light emerging from the polarizer has the *same frequency* as the incident light.

So far so good. But now, let us try to understand this result by thinking about the photons that make up the incident light. The premise here is that all photons in the incident beam are identical. Moreover the photons do not interact with each other. We could even imagine sending the whole energy of the incident light beam one photon at a time. Since all the light that emerges from the polarizer has the same frequency as the incident light, and thus the same frequency, we must conclude that each individual photon either goes through or is absorbed. If a fraction of a photon went through it would be a photon of lower energy and thus lower frequency, which is something that does not happen.

But now we have a problem. As we know from the wave analysis, roughly a fraction $(\cos \alpha)^2$ of the photons must go through, since that is the fraction of the energy that is transmitted. Consequently a fraction $1 - (\cos \alpha)^2$ of the photons must be absorbed. But if all the photons are identical, why is it that what happens to one photon does not happen to all of them?

The answer in quantum mechanics is that there is indeed a loss of determinism. No one can predict if a photon will go through or will get absorbed. The best anyone can do is to predict probabilities. In this case there would be a probability $(\cos \alpha)^2$ of going through and a probability $1 - (\cos \alpha)^2$ of failing to go through.

Two escape routes suggest themselves. Perhaps the polarizer is not really a homogeneous object and depending exactly on where the photon hits it either gets absorbed or goes through. Experiments show this is not the case. A more intriguing possibility was suggested by Einstein and others. A possible way out, they claimed, was the existence of *hidden variables*. The photons, while apparently identical, would have other *hidden* properties, not currently understood, that would determine with certainty which photon goes through and which photon gets absorbed. Hidden variable theories would seem to be untestable, but surprisingly they can be tested. Through the work of John Bell and others, physicists have devised clever experiments that rule out most versions of hidden variable theories. No one has figured out how to restore determinism to quantum mechanics. It seems to be an impossible task.

When we try to describe photons quantum mechanically we could use wavefunctions, or equivalently the language of states. A photon polarized along the $\hat{\mathbf{x}}$ direction is not represented using an electric field, but rather we just give a name for its *state*:

$$|\text{photon}; x\rangle. \quad (3.4)$$

We will learn the rules needed to manipulate such objects, but for the time being you could think of it like a vector in some space yet to be defined. Another state of a photon, or vector is

$$|\text{photon}; y\rangle, \quad (3.5)$$

representing a photon polarized along $\hat{\mathbf{y}}$. These states are the wavefunctions that represent the photon. We now claim that the photons in the beam that is polarized along the direction α are in a state $|\text{photon}; \alpha\rangle$ that can be written as a superposition of the above two states:

$$|\text{photon}; \alpha\rangle = \cos \alpha |\text{photon}; x\rangle + \sin \alpha |\text{photon}; y\rangle. \quad (3.6)$$

This equation should be compared with (3.2). While there are some similarities –both are superpositions– one refers to electric fields and the other to “states” of a single photon. Any photon that emerges from the polarizer will necessarily be polarized in the $\hat{\mathbf{x}}$ direction and therefore it will be in the state

$$\text{Beyond the polarizer: } |\text{photon}; x\rangle. \quad (3.7)$$

This can be compared with (3.3) which with the factor $\cos \alpha$ carries information about the amplitude of the wave. Here, for a single photon, there is no room for such a factor.

In the famous Fifth Solvay International Conference of 1927 the world’s most notable physicists gathered to discuss the newly formulated quantum theory. Seventeen out of the twenty nine attendees were or became Nobel Prize winners. Einstein, unhappy with the uncertainty in quantum mechanics stated the nowadays famous quote: “God does not play dice”, to which Niels Bohr is said to have answered: “Einstein, stop telling God what to do.” Bohr was willing to accept the loss of determinism, Einstein was not.

4 Quantum Superpositions

We have already discussed the concept of linearity; the idea that the sum of two solutions representing physical realities represents a new, allowed, physical reality. This superposition of solutions has a straightforward meaning in classical physics. In the case of electromagnetism, for example, if we have two solutions, each with its own electric and magnetic field, the “sum” solution is simply understood: its electric field is the sum of the electric fields of the two solutions and its magnetic field is the sum of the magnetic fields of the two solutions. In quantum mechanics, as we have explained, linearity holds. The interpretation of a superposition, however, is very surprising.

One interesting example is provided by a Mach-Zehnder interferometer; an arrangement of beam splitters, mirrors, and detectors used by Ernst Mach and Ludwig Zehnder in the 1890’s to study interference between two beams of light.

A beam splitter, as its name indicates, splits an incident beam into two beams, one that is reflected from the splitter and one that goes through the splitter. Our beam-splitters will be balanced: they split a given beam into two beams of equal intensity (Figure 3). The light that bounces off is called the reflected beam, the light that goes through is called the transmitted beam. The incident beam can hit the beam splitter from the top or from the bottom.

The Mach-Zehnder configuration, shown in Figure 4, has a left beam splitter (BS1) and a right beam splitter (BS2). In between we have the two mirrors, M1 on the top and M2 on the bottom. An incoming beam from the left is split by BS1 into two beams, each of which hits a mirror and is then sent into BS2. At BS2 the beams are recombined and sent into two outgoing beams that go into photon detectors D0 and D1.

It is relatively simple to arrange the beam-splitters so that the incoming beam, upon splitting at BS1 and recombination at BS2 emerges in the top beam which goes into D0. In this arrangement no light at all goes into D1. This requires a precise interference effect at BS2. Note that we have

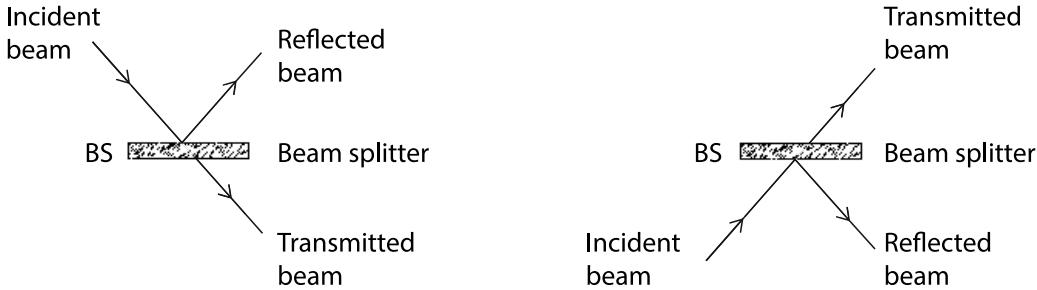


Figure 3: An incident beam hitting a beam-splitter results in a reflected beam and a transmitted beam. Left: incident beam coming from the top. Right: incident beam coming from the bottom.

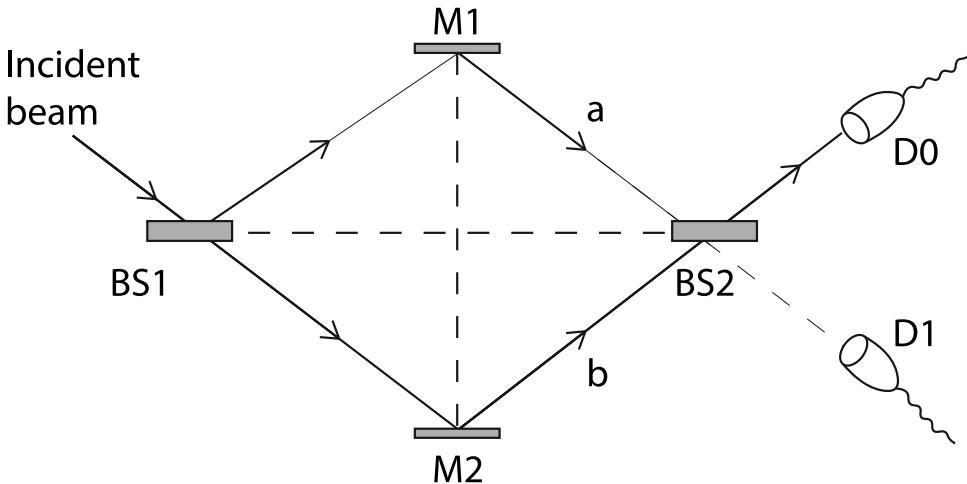


Figure 4: A Mach-Zehnder interferometer consists of two beam splitters BS1 and BS2, two mirrors M1 and M2, and two detectors D0 and D1. An incident beam will be split into two beams by BS1. One beam goes through the upper branch, which contains M1, the other beam goes through the lower branch, which contains M2. The beams on the two branches recombine at BS2 and are then sent into the detectors. The configuration is prepared to produce an interference so that all incident photons end at the detector D0, with none at D1.

two beams incident upon BS2; the top beam is called ‘a’ and the lower beam is called ‘b’. Two contributions go towards D0: the reflection of ‘a’ at BS2 and the transmission from ‘b’ at BS2. These two contributions interfere constructively to give a beam going into D0. Two contributions also go towards D1: the transmission from ‘a’ at BS2 and the reflection from ‘b’ at BS2. These two can indeed be arranged to interfere destructively to give no beam going into D1.

It is instructive to think of the incoming beam as a sequence of photons that we send into the interferometer, one photon at a time. This shows that, at the level of photons, the interference is not interference of one photon with another photon. Each photon must interfere with *itself* to give the result. Indeed interference between two photons is not possible: destructive interference, for example, would require that two photons end up giving no photon, which is impossible by energy conservation.

Therefore, each photon does the very strange thing of going through both branches of the interferometer! Each photon is in a superposition of two states: a state in which the photon is in the top beam or upper branch, added to a state in which the photon is in the bottom beam or lower branch. Thus the state of the photon in the interferometer is a funny state in which the photon seems to be doing two incompatible things at the same time.

Equation (3.6) is another example of a quantum superposition. The photon state has a component along an x -polarized photon and a component along a y -polarized photon.

When we speak of a wavefunction, we also sometimes call it a state, because the wavefunction specifies the “state” of our quantum system. We also sometimes refer to states as vectors. A quantum state may not be a vector like the familiar vectors in three-dimensional space but it is a vector nonetheless because it makes sense to add states and to multiply states by numbers. Just like vectors can be added, linearity guarantees that adding wavefunctions or states is a sensible thing to do. Just like any vector can be written as a sum of other vectors in many different ways, we will do the same with our states. By writing our physical state as sums of other states we can learn about the properties of our state.

Consider now two states $|A\rangle$ and $|B\rangle$. Assume, in addition, that when measuring some property Q in the state $|A\rangle$ the answer is always a , and when measuring the same property Q in the state $|B\rangle$ the answer is always b . Suppose now that our physical state $|\Psi\rangle$ is the superposition

$$|\Psi\rangle = \alpha|A\rangle + \beta|B\rangle, \quad \alpha, \beta \in \mathbb{C}. \quad (4.1)$$

What happens now if we measure property Q in the system described by the state $|\Psi\rangle$? It may seem reasonable that one gets some intermediate value between a and b , but this is not what happens. A measurement of Q will yield either a or b . There is no certain answer, classical determinism is lost, but the answer is always one of these two values and not an intermediate one. The coefficients α and β in the above superposition affect the probabilities with which we may obtain the two possible values. In fact, the probabilities to obtain a or b

$$\text{Probability}(a) \sim |\alpha|^2, \quad \text{Probability}(b) \sim |\beta|^2. \quad (4.2)$$

Since the only two possibilities are to measure a or b , the actual probabilities must sum to one and therefore they are given by

$$\text{Probability}(a) = \frac{|\alpha|^2}{|\alpha|^2 + |\beta|^2}, \quad \text{Probability}(b) = \frac{|\beta|^2}{|\alpha|^2 + |\beta|^2}. \quad (4.3)$$

If we obtain the value a , immediate repeated measurements would still give a , so the state after the measurement must be $|A\rangle$. The same happens for b , so we have

$$\begin{aligned} \text{After measuring } a \text{ the state becomes } |\Psi\rangle &= |A\rangle, \\ \text{After measuring } b \text{ the state becomes } |\Psi\rangle &= |B\rangle. \end{aligned} \quad (4.4)$$

In quantum mechanics one makes the following assumption: *Superposing a state with itself doesn't change the physics*, nor does it change the state in a non-trivial way. Since superimposing a state with itself simply changes the overall number multiplying it, we have that Ψ and $\alpha\Psi$ represent the same physics for any complex number α different from zero. Thus, letting \cong represent physical equivalence

$$|A\rangle \cong 2|A\rangle \cong i|A\rangle \cong -|A\rangle. \quad (4.5)$$

This assumption is necessary to verify that the polarization of a photon state has the expected number of degrees of freedom. The polarization of a plane wave, as one studies in electromagnetism, is described by two real numbers. For this consider an elliptically polarized wave, as shown in Figure 5. At any given point, the electric field vector traces an ellipse whose shape is encoded by the ratio a/b of

the semi-major axes (the first real parameter) and a tilt encoded by the angle θ (the second real parameter). Consider for this a general photon state formed by superposition of the two independent polarization states $|\text{photon}; x\rangle$ and $|\text{photon}; y\rangle$:

$$\alpha|\text{photon}; x\rangle + \beta|\text{photon}; y\rangle, \quad \alpha, \beta \in \mathbb{C}. \quad (4.6)$$

At first sight it looks as if we have two complex parameters α and β , or equivalently, four real parameters. But since the overall factor does not matter we can multiply this state by $1/\alpha$ to get the equivalent state that encodes all the physics

$$|\text{photon}; x\rangle + \frac{\beta}{\alpha}|\text{photon}; y\rangle, \quad (4.7)$$

showing that we really have one complex parameter, the ratio β/α . This is equivalent to two real parameters, as expected.

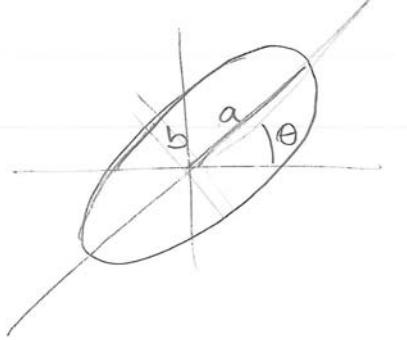


Figure 5: Parameters that define an elliptically polarized state.

Let us do a further example of superposition using electrons. Electrons are particles with spin. Classically, we imagine them as tiny balls spinning around an axis that goes through the particle itself. Once an axis is fixed, the electron has two and only two options: its rotation may be clockwise or counterclockwise about the axis, but in both cases it spins at the same fixed rate. These opposite ways of spinning are called *spin up* and *spin down* along the axis (see Figure 6). The up and down refer to the direction of the angular momentum associated with the rotation, and it is indicated by an arrow. According to quantum mechanics, and as verified by multiple experiments, the same possibilities, up or down, arise *whatever* axis we use to measure the spin of the electron.

Physicists usually set up coordinate systems in space by choosing three orthogonal directions, the directions of the x , y , and z axes. Let us choose to describe our spinning electrons using the z axis. One possible state of an electron is to be spin up along the z axis. Such a state is described as $|\uparrow; z\rangle$, with an arrow pointing up, and the label z indicating that the spin arrow points along the increasing z direction. Another possible state of an electron is spin down along the z axis. Such a state is described as $|\downarrow; z\rangle$, with an arrow pointing down, meaning this time that the spin points along the decreasing z direction. If these two are possible realities, so it would be the state $|\Psi\rangle$ representing the sum

$$|\Psi\rangle = |\uparrow; z\rangle + |\downarrow; z\rangle.$$

The state $|\Psi\rangle$ is in a superposition of a spin up and a spin down state. What kind of physics does this sum $|\Psi\rangle$ represent? It represents a state in which a measurement of the spin along the z axis would result in two possible outcomes with equal probabilities: an electron with spin up or an electron with spin down. Since we can only speak of probabilities, any experiment must involve repetition until

probabilities can be determined. Suppose we had a large ensemble of such electrons, all of them in the above state $|\Psi\rangle$. As we measured their spin along z , one at a time, we would find about half of them spinning up along z and the other half spinning down along z . There is no way to predict which option will be realized as we measure each electron. It is not easy to imagine superposition, but one may try as follows. An electron in the above state is in a different kind of existence in which it is able to both be spinning up along z and spinning down along z simultaneously! It is in such a ghostly, eerie state, doing incompatible things simultaneously, until its spin is measured. Once measured, the electron must immediately choose one of the two options; we always find electrons either spinning up or spinning down.

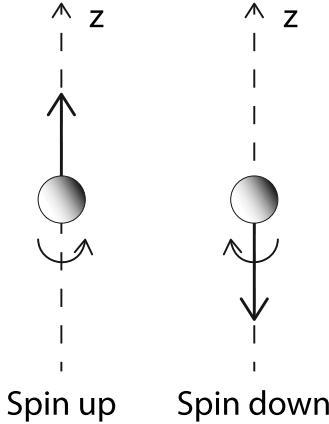


Figure 6: An electron with spin along the z axis. Left: the electron is said to have spin up along z . Right: the electron is said to have spin down along z . The up and down arrows represent the direction of the angular momentum associated with the spinning electron.

A critic of quantum mechanics could suggest a simpler explanation for the above observations. He or she would claim that the following simpler ensemble results in identical experimental results. In the critic's ensemble we have a large number of electrons with 50% of them in the state $|\uparrow; z\rangle$ and 50% of them in the state $|\downarrow; z\rangle$. He or she would then state, correctly, that such an ensemble would yield the same measurements of spins along z as the ensemble of those esoteric $|\Psi\rangle$ states. The new ensemble could provide a simpler explanation of the result without having to invoke quantum superpositions.

Quantum mechanics, however, allows for further experiments that can distinguish between the ensemble of our friendly critic and the ensemble of $|\Psi\rangle$ states. While it would take us too far afield to explain this, if we measured the spin of the electrons in the x direction, instead of z direction, the results would be *different* in the two ensembles. In the ensemble of our critic we would find 50% of the electrons up along x and 50% of the electrons down along x . In our ensemble of $|\Psi\rangle$ states, however, we would find a very simple result: all states pointing up along x . The critic's ensemble is not equivalent to our quantum mechanical ensemble. The critic is thus shown wrong in his or her attempt to show that quantum mechanical superpositions are not required.

5 Entanglement

When we consider superposition of states of *two* particles we can get the remarkable phenomenon called *quantum mechanical entanglement*. Entangled states of two particles are those in which we can't speak separately of the state of each particle. The particles are bound together in a common

state in which they are *entangled* with each other.

Let us consider two non-interacting particles. Particle 1 could be in any of the states

$$\{|u_1\rangle, |u_2\rangle, \dots\}, \quad (5.1)$$

while particle 2 could be in any of the states

$$\{|v_1\rangle, |v_2\rangle, \dots\} \quad (5.2)$$

It may seem reasonable to conclude that the state of the full system, including particle 1 and particle 2 would be specified by stating the state of particle 1 and the state of particle 2. If that would be the case the possible states would be written as

$$|u_i\rangle \otimes |v_j\rangle, \quad i, j \in \mathbb{N}, \quad (5.3)$$

for some specific choice of i and j that specify the state of particle one and particle two, respectively. Here we have used the symbol \otimes , which means *tensor* product, to combine the two states into a single state for the whole system. We will study \otimes later, but for the time being we can think of it as a kind of product that distributes over addition and obeys simple rules, as follows

$$\begin{aligned} (\alpha_1|u_1\rangle + \alpha_2|u_2\rangle) \otimes (\beta_1|v_1\rangle + \beta_2|v_2\rangle) = & \alpha_1\beta_1|u_1\rangle \otimes |v_1\rangle + \alpha_1\beta_2|u_1\rangle \otimes |v_2\rangle \\ & + \alpha_2\beta_1|u_2\rangle \otimes |v_1\rangle + \alpha_2\beta_2|u_2\rangle \otimes |v_2\rangle. \end{aligned} \quad (5.4)$$

The numbers can be moved across the \otimes but the order of the states must be preserved. The state on the left-hand side –expanded out on the right-hand side– is still of the type where we combine a state of the first particle ($\alpha_1|u_1\rangle + \alpha_2|u_2\rangle$) with a state of the second particle ($\beta_1|v_1\rangle + \beta_2|v_2\rangle$). Just like any one of the states listed in (5.3) this state is not entangled.

Using the states in (5.3), however, we can construct more intriguing superpositions. Consider the following one

$$|u_1\rangle \otimes |v_1\rangle + |u_2\rangle \otimes |v_2\rangle. \quad (5.5)$$

A state of two particles is said to be **entangled** if it cannot be written in the factorized form $(\dots) \otimes (\dots)$ which allows us to describe the state by simply stating the state of each particle. We can easily see that the state (5.5) cannot be factorized. If it could it would have to be with a product as indicated in (5.4). Clearly, involving states like $|u_3\rangle$ or $|v_3\rangle$ that do not appear in (5.5) would not help. To determine the constants $\alpha_1, \alpha_2, \beta_1, \beta_2$ we compare the right hand side of (5.4) with our state and conclude that we need

$$\alpha_1\beta_1 = 1, \quad \alpha_1\beta_2 = 0, \quad \alpha_2\beta_1 = 0, \quad \alpha_2\beta_2 = 1. \quad (5.6)$$

It is clear that there is no solution here. The second equation, for example, requires either α_1 or β_2 to be zero. Having $\alpha_1 = 0$ contradicts the first equation, and having $\beta_2 = 0$ contradicts the last equation. This confirms that the state (5.5) is indeed an entangled state. There is no way to describe the state by specifying a state for each of the particles.

Let us illustrate the above discussion using electrons and their spin states. Consider a state of two electrons denoted as $|\uparrow\rangle \otimes |\downarrow\rangle$. As the notation indicates, the first electron, described by the first arrow, is up along z while the second electron, described by the second arrow, is down along z (we omit the label z on the state for brevity). This is not an entangled state. Another possible state is

one where they are doing exactly the opposite: in $|\downarrow\rangle \otimes |\uparrow\rangle$ the first electron is down and the second is up. This second state is also not entangled. It now follows that by superposition we can consider the state

$$|\uparrow\rangle \otimes |\downarrow\rangle + |\downarrow\rangle \otimes |\uparrow\rangle. \quad (5.7)$$

This is an entangled state of the pair of electrons.

Exercise. Show that the above state cannot be factorized and thus is indeed entangled.

In the state (5.7) the first electron is up along z if the second electron is down along z (first term), or the first electron is down along z if the second electron is up along z (second term). There is a correlation between the spins of the two particles; they always point in opposite directions. Imagine that the two entangled electrons are very far away from each other: Alice has one electron of the pair on planet earth and Bob has the other electron on the moon. Nothing we know is connecting these particles but nevertheless the states of the electrons are linked. Measurements we do on the separate particles exhibit correlations. Suppose Alice measures the spin of the electron on earth. If she finds it up along z , it means that the first summand in the above superposition is realized, because in that summand the first particle is up. As discussed before, the state of the two particles immediately becomes that of the first summand. This means that the electron on the moon will *instantaneously* go into the spin down-along- z configuration, something that could be confirmed by Bob, who is sitting in the moon with that particle in his lab. This effect on Bob's electron happens before a message, carried with the speed of light, could reach the moon telling him that a measurement has been done by Alice on the earth particle and the result was spin up. Of course, experiments must be done with an ensemble that contains many pairs of particles, each pair in the same entangled state above. Half of the times the electron on earth will be found up, with the electron on the moon down and the other half of the times the electron on earth will be found down, with the electron on the moon up.

Our friendly critic could now say, correctly, that such correlations between the measurements of spins along z could have been produced by preparing a *conventional* ensemble in which 50% of the pairs are in the state $|\uparrow\rangle \otimes |\downarrow\rangle$ and the other 50% of the pairs are in the state $|\downarrow\rangle \otimes |\uparrow\rangle$. Such objections were dealt with conclusively in 1964 by John Bell, who showed that if Alice and Bob are able to measure spin in *three* arbitrary directions, the correlations predicted by the quantum entangled state are different from the classical correlations of *any* conceivable conventional ensemble. Quantum correlations in entangled states are very subtle and it takes sophisticated experiments to show they are not reproducible as classical correlations. Indeed, experiments with entangled states have confirmed the existence of quantum correlations. The kind of instantaneous action at a distance associated with measurements on well-separated entangled particles does not lead to paradoxes nor, as it may seem, to contradictions with the ideas of special relativity. You cannot use quantum mechanical entangled states to send information faster than the speed of light.

Sarah Geller transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare

<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

SPIN ONE-HALF, BRAS, KETS, AND OPERATORS

B. Zwiebach

September 17, 2013

Contents

1 The Stern-Gerlach Experiment	1
2 Spin one-half states and operators	7
3 Properties of Pauli matrices and index notation	12
4 Spin states in arbitrary direction	16

1 The Stern-Gerlach Experiment

In 1922, at the University of Frankfurt (Germany), Otto Stern and Walther Gerlach, did fundamental experiments in which beams of silver atoms were sent through inhomogeneous magnetic fields to observe their deflection. These experiments demonstrated that these atoms have quantized magnetic moments that can take two values. Although consistent with the idea that the electron had spin, this suggestion took a few more years to develop.

Pauli introduced a “two-valued” degree of freedom for electrons, without suggesting a physical interpretation. Kronig suggested in 1925 that it this degree of freedom originated from the self-rotation of the electron. This idea was severely criticized by Pauli, and Kronig did not publish it. In the same year Uhlenbeck and Goudsmit had a similar idea, and Ehrenfest encouraged them to publish it. They are presently credited with the discovery that the electron has an intrinsic spin with value “one-half”. Much of the mathematics of spin one-half was developed by Pauli himself in 1927. It took in fact until 1927 before it was realized that the Stern-Gerlach experiment did measure the magnetic moment of the electron.

A current on a closed loop induces a magnetic dipole moment. The magnetic moment vector $\vec{\mu}$ is proportional to the current I on the loop and the area A of the loop:

$$\vec{\mu} = IA \vec{A}. \quad (1.1)$$

The vector area, for a planar loop is a vector normal to the loop and of length equal to the value of the area. The direction of the normal is determined from the direction of the current and the right-hand rule. The product μB of the magnetic moment times the magnetic field has units of energy, thus the units of μ are

$$[\mu] = \frac{\text{erg}}{\text{gauss}} \text{ or } \frac{\text{Joule}}{\text{Tesla}} \quad (1.2)$$

When we have a charge distribution spinning we get a magnetic moment and, if the distribution has mass, an angular momentum. The magnetic moment and the angular momentum are proportional to each other, and the constant of proportionality is universal. To see this consider rotating radius R ring of charge with uniform charge distribution and total charge Q . Assume the ring is rotating about an axis perpendicular to the plane of the ring and going through its center. Let the tangential velocity at the ring be v . The current at the loop is equal to the linear charge density λ times the velocity:

$$I = \lambda v = \frac{Q}{2\pi R} v. \quad (1.3)$$

It follows that the magnitude μ of the dipole moment of the loop is

$$\mu = IA = \frac{Q}{2\pi R} v \pi R^2 = \frac{Q}{2} Rv. \quad (1.4)$$

Let the mass of the ring be M . The magnitude L of the angular momentum of the ring is then $L = R(Mv)$. As a result

$$\mu = \frac{Q}{2M} RMv = \frac{Q}{2M} L, \quad (1.5)$$

leading to the notable ratio

$$\frac{\mu}{L} = \frac{Q}{2M}.$$

(1.6)

Note that the ratio does not depend on the radius of the ring, nor on its velocity. By superposition, any rotating distribution with uniform mass and charge density will have a ratio μ/L as above, with Q the total charge and M the total mass. The above is also written as

$$\mu = \frac{Q}{2M} L. \quad (1.7)$$

an classical electron going in a circular orbit around a nucleus will have both orbital angular momentum and a magnetic moment, related as above, with Q the electron charge and M the electron mass. In quantum mechanics the electron is not actually going in circles around the proton, but the right quantum analog exhibits both orbital angular momentum and magnetic moment.

We can ask if the electron can have an intrinsic μ , as if it were, a tiny spinning ball. Well, it has an intrinsic μ but it cannot really be viewed as a rotating little ball of charge (this was part of Pauli's objection to the original idea of spin). Moreover, we currently view the electron as an elementary particle with zero size, so the idea that it rotates is just not sensible. The classical relation, however, points to the correct result. Even if it has no size, the electron has an intrinsic spin S -intrinsic angular momentum. One could guess that

$$\mu = \frac{e}{2m_e} S ? \quad (1.8)$$

Since angular momentum and spin have the same units we write this as

$$\mu = \frac{e\hbar}{2m_e} \frac{S}{\hbar} ? \quad (1.9)$$

This is not exactly right. For electrons the magnetic moment is twice as large as the above relation suggests. One uses a constant “ g -factor” to describe this

$$\mu = g \frac{e\hbar}{2m_e} \frac{S}{\hbar}, \quad g = 2 \text{ for an electron.} \quad (1.10)$$

This factor of two is in fact predicted by the Dirac equation for the electron, and has been verified experimentally. To describe the above more briefly, one introduces the canonical value μ_B of the dipole moment called the Bohr-magneton:

$$\mu_B \equiv \frac{e\hbar}{2m_e} = 9.27 \times 10^{-24} \frac{\text{J}}{\text{Tesla}}. \quad (1.11)$$

With this formula we get

$$\mu = g \mu_B \frac{S}{\hbar}, \quad g = 2 \text{ for an electron.} \quad (1.12)$$

Both the magnetic moment and the angular momentum point in the same direction if the charge is positive. For the electron we thus get

$$\vec{\mu} = -g \mu_B \frac{\vec{S}}{\hbar}, \quad g = 2. \quad (1.13)$$

Another feature of magnetic dipoles is needed for our discussion. A dipole placed in a non-uniform magnetic field will experience a force. An illustration is given in Figure 1 below, where to the left we show a current ring whose associated dipole moment $\vec{\mu}$ points upward. The magnetic field lines diverge as we move up, so the magnetic field is stronger as we move down. This dipole will experience a force pointing down, as can be deduced by basic considerations. On a small piece of wire the force $d\vec{F}$ is proportional to $\vec{I} \times \vec{B}$. The vectors $d\vec{F}$ are sketched in the right part of the figure. Their horizontal components cancel out, but the result is a net force downwards.

In general the equation for the force on a dipole $\vec{\mu}$ in a magnetic field \vec{B} is given by

$$\vec{F} = \nabla(\vec{\mu} \cdot \vec{B}). \quad (1.14)$$

Note that the force points in the direction for which $\vec{\mu} \cdot \vec{B}$ increases the fastest. Given that in our situation $\vec{\mu}$ and \vec{B} are parallel, this direction is the direction in which the magnitude of \vec{B} increases the fastest.

The Stern-Gerlach experiment uses atoms of silver. Silver atoms have 47 electrons. Forty-six of them fill completely the $n = 1, 2, 3$, and 4 levels. The last electron is an $n = 5$ electron with zero orbital angular momentum (a $5s$ state). The only possible angular momentum is the intrinsic angular momentum of the last electron. Thus a magnetic dipole moment is also that of the last electron (the nucleus has much smaller dipole moment and can be ignored). The silver is vaporized in an oven and with a help of a collimating slit a narrow beam of silver atoms is send down to a magnet configuration.

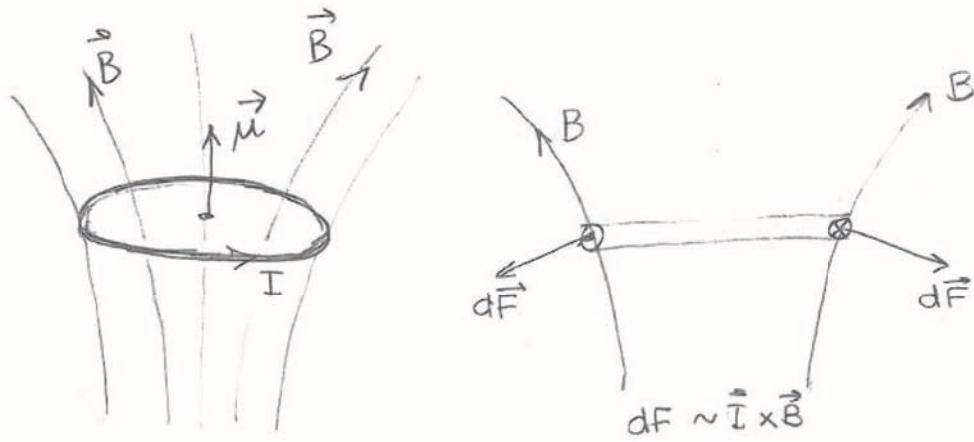


Figure 1: A magnetic dipole in a non-uniform magnetic field will experience a force. The force points in the direction for which $\vec{\mu} \cdot \vec{B}$ grows the fastest. In this case the force is downward.

In the situation described by Figure 2 the magnetic field points mostly in the positive z direction, and the gradient is also in the positive z -direction. As a result, the above equation gives

$$\vec{F} \simeq \nabla(\mu_z B_z) = \mu_z \nabla B_z \simeq \mu_z \frac{\partial B_z}{\partial z} \vec{e}_z, \quad (1.15)$$

and the atoms experience a force in the z -direction proportional to the z -component of their magnetic moment. Undeflected atoms would hit the detector screen at the point P . Atoms with positive μ_z should be deflected upwards and atoms with negative μ_z should be deflected downwards.

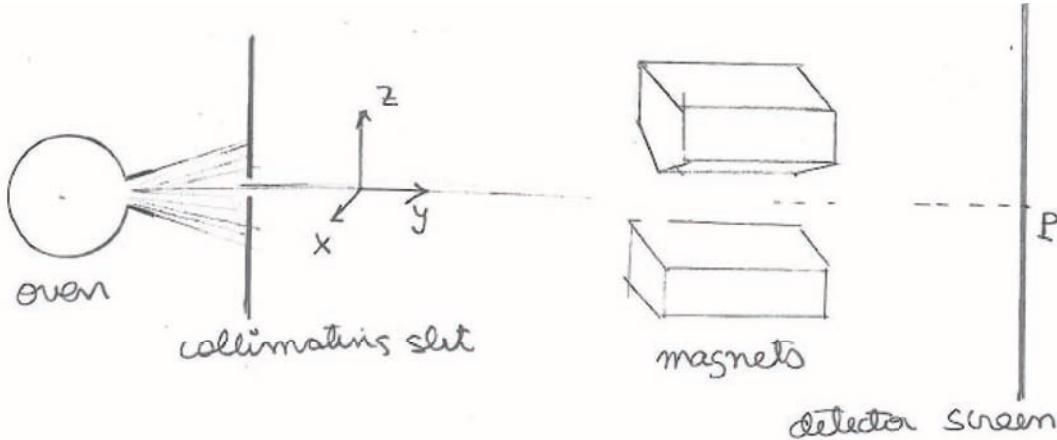


Figure 2: A sketch of the Stern-Gerlach apparatus. An oven and a collimating slit produces a narrow beam of silver atoms. The beam goes through a region with a strong magnetic field and a strong gradient, both in the z -direction. A screen, to the right, acts as a detector.

The oven source produces atoms with magnetic moments pointing in random directions and thus

the expectation was that the z-component of the magnetic moment would define a smooth probability distribution leading to a detection that would be roughly like the one indicated on the left side of Figure 3. Surprisingly, the observed result was two separate peaks as if all atoms had either a fixed positive μ_z or a fixed negative μ_z . This is shown on the right side of the figure. The fact that the peaks are spatially separated led to the original cumbersome name of “space quantization.” The Stern Gerlach experiment demonstrates the quantization of the dipole moment, and by theoretical inference from (1.13), the quantization of the spin (or intrinsic) angular momentum.

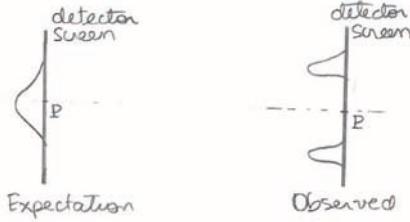


Figure 3: Left: the pattern on the detector screen that would be expected from classical physics. Right: the observed pattern, showing two separated peaks corresponding to up and down magnetic moments.

It follows from (1.13) that

$$\mu_z = -2\mu_B \frac{S_z}{\hbar}. \quad (1.16)$$

The deflections calculated using the details of the magnetic field configuration are consistent with

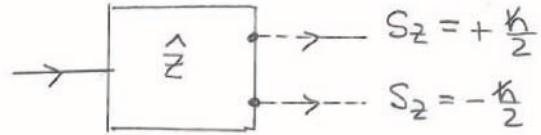
$$S_z = \pm \frac{\hbar}{2}, \text{ or } \frac{S_z}{\hbar} = \pm \frac{1}{2}. \quad (1.17)$$

A particle with such possible values of S_z/\hbar is called a spin one-half particle. The magnitude of the magnetic moments is one Bohr magneton.

With the magnetic field and its gradient along the z-direction, the Stern-Gerlach apparatus measures the component of the spin \vec{S} in the z direction. To streamline our pictures we will denote such apparatus as a box with a \hat{z} label, as in Figure 4. The box lets the input beam come in from the left and lets out two beams from the right side. If we placed a detector to the right, the top beam would be identified as having atoms with $S_z = \hbar/2$ and the bottom having atoms with $S_z = -\hbar/2$.¹

Let us now consider thought experiments in which we put a few SG apparatus in series. In the first configuration, shown at the top of Figure 5, the first box is a \hat{z} SG machine, where we block the $S_z = -\hbar/2$ output beam and let only the $S_z = \hbar/2$ beam go into the next machine. This machine acts as a filter. The second SG apparatus is also a \hat{z} machine. Since all ingoing particles have $S_z = \hbar/2$ the second machine lets those out the top output and nothing comes out in the bottom output. The quantum mechanical lesson here is that $S_z = \hbar/2$ states have no component or amplitude along $S_z = -\hbar/2$. These are said to be orthogonal states.

¹In the quantum mechanical view of the experiment, a single atom can be in both beams, with different amplitudes. Only the act of measurement, which corresponds to the act of placing the detector screen, forces the atom to decide in which beam it is.



SG machine measuring S_z

Figure 4: Left: A schematic representation of the SG apparatus, minus the screen.

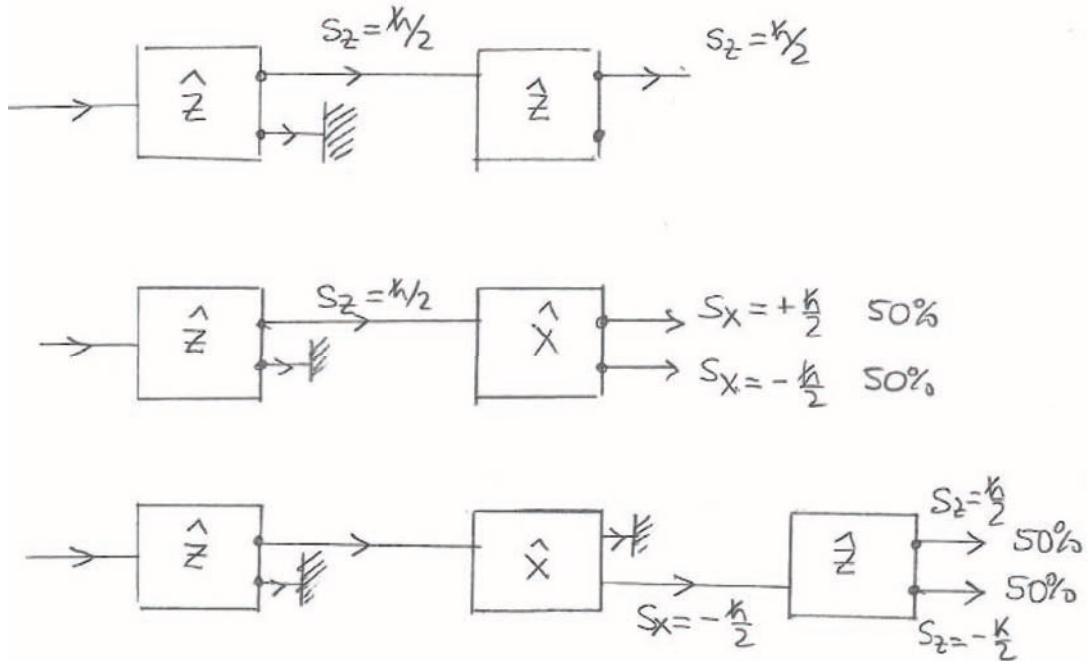


Figure 5: Left: Three configurations of SG boxes.

The second configuration in the figure shows the outgoing $S_z = \hbar/2$ beam from the first machine going into an \hat{x} -machine. The outputs of this machine are –in analogy to the \hat{z} machine– $S_x = \hbar/2$ and $S_x = -\hbar/2$. Classically an object with angular momentum along the z axis has no component of angular momentum along the x axis, these are orthogonal directions. But the result of the experiment indicates that quantum mechanically this is not true for spins. About half of the $S_z = \hbar/2$ atoms exit through the top $S_x = \hbar/2$ output, and the other half exit through the bottom $S_x = -\hbar/2$ output. Quantum mechanically, a state with a definite value of S_z has an amplitude along the state $S_x = \hbar/2$ as well as an amplitude along the state $S_x = -\hbar/2$.

In the third and bottom configuration the $S_z = \hbar/2$ beam from the first machine goes into the \hat{x} machine and the top output is blocked so that we only have an $S_x = -\hbar/2$ output. That beam is

fed into a \hat{z} type machine. One could speculate that the beam entering the third machine has both $S_x = -\hbar/2$ and $S_z = \hbar/2$, as it is composed of silver atoms that made it through both machines. If that were the case the third machine would let all atoms out the top output. This speculation is falsified by the result. There is no memory of the first filter: the particles out of the second machine do not anymore have $S_z = \hbar/2$. We find half of the particles make it out of the third machine with $S_z = \hbar/2$ and the other half with $S_z = -\hbar/2$. In the following section we discuss a mathematical framework consistent with the results of the above thought experiments.

2 Spin one-half states and operators

The SG experiment suggests that the spin states of the electron can be described using two basis vectors (or kets):

$$|z;+\rangle \quad \text{and} \quad |z;-\rangle. \quad (2.1)$$

The first corresponds to an electron with $S_z = \frac{\hbar}{2}$. The z label indicates the component of the spin, and the + the fact that the component of spin is positive. This state is also called ‘spin up’ along z . The second state corresponds to an electron with $S_z = -\frac{\hbar}{2}$, that is a ‘spin down’ along z . Mathematically, we have an operator \hat{S}_z for which the above states are eigenstates, with opposite eigenvalues:

$$\begin{aligned} \hat{S}_z |z;+\rangle &= +\frac{\hbar}{2} |z;+\rangle \\ \hat{S}_z |z;-\rangle &= -\frac{\hbar}{2} |z;-\rangle. \end{aligned} \quad (2.2)$$

If we have two basis states, then the state space of electron spin is a two-dimensional *complex vector space*. Each vector in this vector space represents a possible state of the electron spin. We are not discussing other degrees of freedom of the electron, such as its position, momentum, or energy. The general vector in the two-dimensional space is an arbitrary linear combination of the basis states and thus takes the form

$$|\Psi\rangle = c_1 |z;+\rangle + c_2 |z;-\rangle, \quad \text{with } c_1, c_2 \in \mathbb{C} \quad (2.3)$$

It is customary to call the state $|z;+\rangle$ the *first* basis state and it denote by $|1\rangle$. The state $|z;-\rangle$ is called the *second* basis state and is denoted by $|2\rangle$. States are vectors in some vector space. In a two-dimensional vector space a vector is explicitly *represented* as a column vector with two components. The first basis vector is represented as $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and the second basis vector is represented as $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Thus we have the following names for states and their concrete representation as column vectors

$$\begin{aligned} |z:+\rangle &= |1\rangle \longleftrightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ |z:-\rangle &= |2\rangle \longleftrightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned} \quad (2.4)$$

Using these options the state in (2.3) takes the possible forms

$$|\Psi\rangle = c_1|z;+\rangle + c_2|z;-\rangle = c_1|1\rangle + c_2|2\rangle \longleftrightarrow c_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \quad (2.5)$$

As we mentioned before, the top experiment in Figure 5 suggests that we have an orthonormal basis. The state $|z;+\rangle$ entering the second machine must have zero overlap with $|z,-\rangle$ since no such down spins emerge. Moreover the overlap of $|z;+\rangle$ with itself must be one, as all states emerge from the second machine top output. We thus write

$$\langle z;-|z;+\rangle = 0, \quad \langle z;+|z;+\rangle = 1. \quad (2.6)$$

and similarly, we expect

$$\langle z;+|z;-\rangle = 0, \quad \langle z;-|z;-\rangle = 1. \quad (2.7)$$

Using the notation where the basis states are labeled as $|1\rangle$ and $|2\rangle$ we have the simpler form that summarizes the four equations above:

$$\langle i|j\rangle = \delta_{ij}, \quad i, j = 1, 2. \quad (2.8)$$

We have not yet made precise what we mean by the ‘bras’ so let us do so briefly. We define the basis ‘bras’ as the *row vectors* obtained by transposition and complex conjugation:

$$\langle 1| \longleftrightarrow (1, 0), \quad \langle 2| \longleftrightarrow (0, 1). \quad (2.9)$$

Given states $|\alpha\rangle$ and $|\beta\rangle$

$$\begin{aligned} |\alpha\rangle &= \alpha_1|1\rangle + \alpha_2|2\rangle \longleftrightarrow \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \\ |\beta\rangle &= \beta_1|1\rangle + \beta_2|2\rangle \longleftrightarrow \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \end{aligned} \quad (2.10)$$

we associate

$$\langle \alpha| \equiv \alpha_1^* \langle 1| + \alpha_2^* \langle 2| \longleftrightarrow (\alpha_1^*, \alpha_2^*) \quad (2.11)$$

and the ‘bra-ket’ inner product is defined as the ‘obvious’ matrix product of the row vector and column vector representatives:

$$\langle \alpha|\beta\rangle \equiv (\alpha_1^*, \alpha_2^*) \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \alpha_1^*\beta_1 + \alpha_2^*\beta_2. \quad (2.12)$$

Note that this definition is consistent with (2.8).

When we represent the states as two-component column vectors the operators that act on the states to give new states can be *represented* as two-by-two matrices. We can thus represent the operator \hat{S}_z as a 2×2 matrix which we claim takes the form

$$\hat{S}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(2.13)

To test this, it suffices to verify that the matrix \hat{S}_z acts on the column vectors that represent the basis states as expected from (2.2). Indeed,

$$\begin{aligned}\hat{S}_z|z;+\rangle &= +\frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = +\frac{\hbar}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = +\frac{\hbar}{2}|z;+\rangle \\ \hat{S}_z|z;-\rangle &= +\frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{\hbar}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{\hbar}{2}|z;-\rangle.\end{aligned}\tag{2.14}$$

In fact, the states $|1\rangle$ and $|2\rangle$, viewed as column vectors are the eigenstates of matrix \hat{S}_z .

There is nothing particular about the z axis. We could have started with a SG apparatus that measures spin along the x axis and we would have been led to an operator \hat{S}_x . Had we used the y axis we would have been led to the operator \hat{S}_y . Since spin represents angular momentum (albeit of intrinsic type), it is expected to have three components, just like orbital angular momentum has three components: \hat{L}_x , \hat{L}_y , and \hat{L}_z . These are all hermitian operators, written as products of coordinates and momenta in three-dimensional space. Writing $\hat{L}_x = \hat{L}_1$, $\hat{L}_y = \hat{L}_2$, and $\hat{L}_z = \hat{L}_3$, their commutation relations can be briefly stated as

$$[\hat{L}_i, \hat{L}_j] = i\hbar\epsilon_{ijk}\hat{L}_k\tag{2.15}$$

This is the famous algebra of angular momentum, repeated indices are summed over the values 1,2,3, and ϵ_{ijk} is the totally antisymmetric symbol with $\epsilon_{123} = +1$. Make sure that you understand this notation clearly, and can use it to see that it implies the relations

$$\begin{aligned}[\hat{L}_x, \hat{L}_y] &= i\hbar\hat{L}_z, \\ [\hat{L}_y, \hat{L}_z] &= i\hbar\hat{L}_x, \\ [\hat{L}_z, \hat{L}_x] &= i\hbar\hat{L}_y.\end{aligned}\tag{2.16}$$

While, for example, $\hat{L}_z = \hat{x}\hat{p}_y - \hat{y}\hat{p}_x$ is a hermitian operator written in terms of coordinates and momenta, we have no such construction for \hat{S}_z . The latter is a more abstract operator, it does not act on wavefunctions $\psi(\vec{x})$ but rather on the 2-component column vectors introduced above. The operator \hat{S}_z is just a two-by-two *hermitian*² matrix with constant entries! If spin is a quantum mechanical angular momentum, we must have that the triplet of operators \hat{S}_z , \hat{S}_x , and \hat{S}_y satisfy

$$\begin{aligned}[\hat{S}_x, \hat{S}_y] &= i\hbar\hat{S}_z, \\ [\hat{S}_y, \hat{S}_z] &= i\hbar\hat{S}_x, \\ [\hat{S}_z, \hat{S}_x] &= i\hbar\hat{S}_y,\end{aligned}\tag{2.17}$$

or, again using numerical subscripts for the components ($\hat{S}_1 = \hat{S}_x, \dots$) we must have

$$[\hat{S}_i, \hat{S}_j] = i\hbar\epsilon_{ijk}\hat{S}_k.\tag{2.18}$$

²Hermitian means that the matrix is preserved by taking the operations of transposition and complex conjugation.

We can now try to figure out how the matrices for \hat{S}_x and \hat{S}_y must look, given that we know the matrix for \hat{S}_z . We have a few constraints. First the matrices must be hermitian, just like the angular momentum operators are. Two-by-two hermitian matrices take the form

$$\begin{pmatrix} 2c & a - ib \\ a + ib & 2d \end{pmatrix}, \text{ with } a, b, c, d \in \mathbb{R} \quad (2.19)$$

Indeed, you can easily see that transposing and complex conjugating gives exactly the same matrix. Since the two-by-two identity matrix commutes with every matrix, we can subtract from the above matrix any multiple of the identity without any loss of generality. Subtracting the identity matrix times $(c + d)$ we find the still hermitian matrix

$$\begin{pmatrix} c - d & a - ib \\ a + ib & d - c \end{pmatrix}, \text{ with } a, b, c, d \in \mathbb{R} \quad (2.20)$$

Since we are on the lookout for \hat{S}_x and \hat{S}_y we can subtract a matrix proportional to \hat{S}_z . Since \hat{S}_z is diagonal with entries of same value but opposite signs, we can cancel the diagonal terms above and are left over with

$$\begin{pmatrix} 0 & a - ib \\ a + ib & 0 \end{pmatrix}, \text{ with } a, b \in \mathbb{R} \quad (2.21)$$

Thinking of the space of two-by-two hermitian matrices as a real vector space, the hermitian matrices given above can be associated to two basis “vectors” that are the matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad (2.22)$$

since multiplying the first by the real constant a and the second by the real constant b and adding gives us the matrix above. In fact, together with the identity matrix and the \hat{S}_z matrix, with the $\hbar/2$ deleted,

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (2.23)$$

we got the complete set of four two-by-two matrices that viewed as basis vectors in a real vector space, can be used to build the most general hermitian two-by-two matrix by using real linear combinations.

Back to our problem, we are supposed to find \hat{S}_x and \hat{S}_y among the matrices in (2.22). The overall scale of the matrices can be fixed by the constraint that their eigenvalues be $\pm\hbar/2$, just like they are for \hat{S}_z . Let us give the eigenvalues (denoted by λ) and the associated normalized eigenvectors for these two matrices. Short computations (can you do them?) give

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} : \quad \lambda = 1, \text{ for } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \lambda = -1, \text{ for } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad (2.24)$$

for the first matrix and

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} : \quad \lambda = 1, \text{ for } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \lambda = -1, \text{ for } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}, \quad (2.25)$$

for the second matrix. In case you are puzzled by the normalizations, note that a vector $\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ is normalized if $|c_1|^2 + |c_2|^2 = 1$. Since the eigenvalues of both matrices are ± 1 , we tentatively identify

$$\hat{S}_x = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{S}_y = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad (2.26)$$

which have, at least, the correct eigenvalues. But in fact, these also satisfy the commutation relations! Indeed, we check that, as desired,

$$\begin{aligned} [\hat{S}_x, \hat{S}_y] &= \frac{\hbar^2}{4} \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} - \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) \\ &= \frac{\hbar^2}{4} \left(\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} - \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \right) \\ &= \frac{\hbar^2}{4} \begin{pmatrix} 2i & 0 \\ 0 & -2i \end{pmatrix} = i\hbar \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = i\hbar \hat{S}_z. \end{aligned} \quad (2.27)$$

All in all we have

$$\boxed{\hat{S}_x = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{S}_y = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{S}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.} \quad (2.28)$$

Exercise. Verify that the above matrices satisfy the other two commutation relations in (2.17).

You could ask if we got the unique solution for \hat{S}_x and \hat{S}_y given the choice of \hat{S}_z ? The answer is no, but it is not our fault, as illustrated by the following check:

Exercise. Check that the set of commutation relations of the spin operators are in fact preserved when we replace $\hat{S}_x \rightarrow -\hat{S}_y$ and $\hat{S}_y \rightarrow \hat{S}_x$.

The solution we gave is the one conventionally used by all physicists. Any other solution is physically equivalent to the one we gave (as will be explained in more detail after we develop more results). The solution defines the **Pauli matrices** σ_i by writing

$$\hat{S}_i = \frac{\hbar}{2} \sigma_i. \quad (2.29)$$

We then have that the Pauli matrices are

$$\boxed{\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.} \quad (2.30)$$

Let us describe the eigenstates of \hat{S}_x , which given (2.26) can be read from (2.24):

$$\hat{S}_x |x; \pm\rangle = \pm |x; \pm\rangle. \quad (2.31)$$

with

$$\boxed{\begin{aligned} |x; +\rangle &= \frac{1}{\sqrt{2}} |z; +\rangle + \frac{1}{\sqrt{2}} |z; -\rangle \longleftrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \\ |x; -\rangle &= \frac{1}{\sqrt{2}} |z; +\rangle - \frac{1}{\sqrt{2}} |z; -\rangle \longleftrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \end{aligned}} \quad (2.32)$$

Note that these states are orthogonal to each other. The above equations can be inverted to find

$$\begin{aligned} |z;+\rangle &= \frac{1}{\sqrt{2}}|x;+\rangle + \frac{1}{\sqrt{2}}|x;-\rangle \\ |z;-\rangle &= \frac{1}{\sqrt{2}}|x;+\rangle - \frac{1}{\sqrt{2}}|x;-\rangle \end{aligned} \quad (2.33)$$

These relations are consistent with the second experiment shown in Figure 5. The state $|z;+\rangle$ entering the second, \hat{x} -type SG apparatus, has equal probability to be found in $|x;+\rangle$ as it has probability to be found in $|x;-\rangle$. This is reflected in the first of the above relations, since we have the amplitudes

$$\langle x;+|z;+\rangle = \frac{1}{\sqrt{2}}, \quad \langle x;-|z;+\rangle = \frac{1}{\sqrt{2}}. \quad (2.34)$$

These probabilities, being equal to the norm squared of the amplitudes, are $1/2$ in both cases. The relative minus sign on the second equation above is needed to make it orthogonal to the state on the first equation.

We can finally consider the eigenstates of \hat{S}_y axis. We have

$$\hat{S}_y|y;\pm\rangle = \pm\frac{\hbar}{2}|y;\pm\rangle. \quad (2.35)$$

and using (2.25) we read

$$\begin{aligned} |y;+\rangle &= \frac{1}{\sqrt{2}}|z;+\rangle + \frac{i}{\sqrt{2}}|z;-\rangle \longleftrightarrow \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}, \\ |y;-\rangle &= \frac{1}{\sqrt{2}}|z;+\rangle - \frac{i}{\sqrt{2}}|z;-\rangle \longleftrightarrow \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}. \end{aligned} \quad (2.36)$$

Note that this time the superposition of $|z;\pm\rangle$ states involves complex numbers (there would be no way to find y type states without them).

3 Properties of Pauli matrices and index notation

Since we know the commutation relations for the spin operators

$$[\hat{S}_i, \hat{S}_j] = i\hbar\epsilon_{ijk}\hat{S}_k, \quad (3.37)$$

and we have $S_i = \frac{\hbar}{2}\sigma_i$, it follows that

$$\frac{\hbar}{2}\frac{\hbar}{2}[\sigma_i, \sigma_j] = i\hbar\epsilon_{ijk}\frac{\hbar}{2}\sigma_k. \quad (3.38)$$

Cancelling the \hbar 's and some factors of two, we find

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k. \quad (3.39)$$

Another important property of the Pauli matrices is that they square to the identity matrix. This is best checked explicitly (do it!):

$$(\sigma_1)^2 = (\sigma_2)^2 = (\sigma_3)^2 = \mathbf{1}. \quad (3.40)$$

This property “explains” that the eigenvalues of each of the Pauli matrices could only be plus or minus one. Indeed, the eigenvalues of a matrix satisfy the algebraic equation that the matrix satisfies. Take for example a matrix M that satisfies the matrix equation

$$M^2 + \alpha M + \beta \mathbf{1} = 0 \quad (3.41)$$

Let v be an eigenvector of M with eigenvalue λ : $Mv = \lambda v$. Let the above equation act on v

$$M^2 v + \alpha M v + \beta \mathbf{1} v = 0 \rightarrow \lambda^2 v + \alpha \lambda v + \beta v = 0 \rightarrow (\lambda^2 + \alpha \lambda + \beta)v = 0, \quad (3.42)$$

and since $v \neq 0$ (by definition an eigenvector cannot be zero!) we conclude that $\lambda^2 + \alpha \lambda + \beta = 0$, as claimed. For the case of the Pauli matrices we have $(\sigma_i)^2 = 1$ and therefore the eigenvalues must satisfy $\lambda^2 = 1$. As a result, $\lambda = \pm 1$ are the only options.

We also note, by inspection, that the Pauli matrices have zero trace, namely, the sum of entries on the diagonal is zero:

$$\text{tr}(\sigma_i) = 0, \quad i = 1, 2, 3. \quad (3.43)$$

A fact from linear algebra is that the trace of a matrix is equal to the sum of its eigenvalues. So each Pauli matrix must have two eigenvalues that add up to zero. Since the eigenvalues can only be plus or minus one, we must have one of each. This shows that each of the Pauli matrices has a plus one and a minus one eigenvalue.

If you compute a commutator of Pauli matrices by hand you might notice a curious property. Take the commutator of σ_1 and σ_2 :

$$[\sigma_1, \sigma_2] = \sigma_1 \sigma_2 - \sigma_2 \sigma_1. \quad (3.44)$$

The two contributions on the right hand side give

$$\begin{aligned} \sigma_1 \sigma_2 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \\ \sigma_2 \sigma_1 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}. \end{aligned} \quad (3.45)$$

The second contribution is minus the first, so that both terms contribute equally to the commutator! In other words,

$$\sigma_1 \sigma_2 = -\sigma_2 \sigma_1. \quad (3.46)$$

This equation is taken to mean that σ_1 and σ_2 *anticommute*. Just like we define the commutator of two operators X, Y by $[X, Y] \equiv XY - YX$, we define the **anticommutator**, denoted by curly brackets, by

$$\text{Anticommutator: } \{X, Y\} \equiv XY + YX. \quad (3.47)$$

In this language we have checked that

$$\{\sigma_1, \sigma_2\} = 0, \quad (3.48)$$

and the property $\sigma_1^2 = \mathbf{1}$, for example, can be rewritten as

$$\{\sigma_1, \sigma_1\} = 2 \cdot \mathbf{1}. \quad (3.49)$$

In fact, as you can check (two cases to examine) that any two different Pauli matrices anticommute:

$$\{\sigma_i, \sigma_j\} = 0, \quad \text{for } i \neq j. \quad (3.50)$$

We can easily improve on this equation to make it work also when i is equal to j . We claim that

$$\boxed{\{\sigma_i, \sigma_j\} = 2\delta_{ij} \mathbf{1}.} \quad (3.51)$$

Indeed, when $i \neq j$ the right-hand side vanishes, as needed, and when i is equal to j , the right-hand side gives $2 \cdot \mathbf{1}$, also as needed in view of (3.49) and its analogs for the other Pauli matrices.

Both the commutator and anti-commutator identities for the Pauli matrices can be summarized in a single equation. This is possible because, for any two operators X, Y we have

$$X Y = \frac{1}{2} \{X, Y\} + \frac{1}{2} [X, Y], \quad (3.52)$$

as you should confirm by expansion. Applied to the product of two Pauli matrices and using our expressions for the commutator and anticommutator we get

$$\boxed{\sigma_i \sigma_j = \delta_{ij} \mathbf{1} + i \epsilon_{ijk} \sigma_k.} \quad (3.53)$$

This equation can be recast in vector notation. Denote by bold symbols three-component vectors, for example, $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$. Then the dot product

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3 = a_i b_i = a_i b_j \delta_{ij}. \quad (3.54)$$

Note the use of the sum convention: repeated indices are summed over. Moreover, note that $b_j \delta_{ij} = b_i$ (can you see why?). We also have that

$$\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2. \quad (3.55)$$

Cross products use the epsilon symbol. Make sure you understand why

$$(\mathbf{a} \times \mathbf{b})_k = a_i b_j \epsilon_{ijk}. \quad (3.56)$$

We can also have triplets of operators, or matrices. For the Pauli matrices we denote

$$\boldsymbol{\sigma} \equiv (\sigma_1, \sigma_2, \sigma_3). \quad (3.57)$$

We can construct a matrix by dot product of a vector \mathbf{a} with the ‘vector’ $\boldsymbol{\sigma}$. We define

$$\mathbf{a} \cdot \boldsymbol{\sigma} \equiv a_1 \sigma_1 + a_2 \sigma_2 + a_3 \sigma_3 = a_i \sigma_i. \quad (3.58)$$

Note that $\mathbf{a} \cdot \boldsymbol{\sigma}$ is just a single two-by-two matrix. Since the components of \mathbf{a} are numbers, and numbers commute with matrices, this dot product is commutative: $\mathbf{a} \cdot \boldsymbol{\sigma} = \boldsymbol{\sigma} \cdot \mathbf{a}$. We are now ready to rewrite (3.53). Multiply this equation by $a_i b_j$ to get

$$\begin{aligned} a_i \sigma_i b_j \sigma_j &= a_i b_j \delta_{ij} \mathbf{1} + i (a_i b_j \epsilon_{ijk}) \sigma_k \\ &= (\mathbf{a} \cdot \mathbf{b}) \mathbf{1} + i (\mathbf{a} \times \mathbf{b})_k \sigma_k, \end{aligned} \quad (3.59)$$

so that, finally, we get the matrix equation

$$(\mathbf{a} \cdot \boldsymbol{\sigma})(\mathbf{b} \cdot \boldsymbol{\sigma}) = (\mathbf{a} \cdot \mathbf{b}) \mathbf{1} + i (\mathbf{a} \times \mathbf{b}) \cdot \boldsymbol{\sigma}. \quad (3.60)$$

As a simple application we take $\mathbf{b} = \mathbf{a}$. We then have $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$ and $\mathbf{a} \times \mathbf{a} = 0$, so that the above equation gives

$$(\mathbf{a} \cdot \boldsymbol{\sigma})^2 = |\mathbf{a}|^2 \mathbf{1}. \quad (3.61)$$

When \mathbf{a} is a unit vector this becomes

$$(\mathbf{n} \cdot \boldsymbol{\sigma})^2 = \mathbf{1}, \quad \mathbf{n} \text{ a unit vector.} \quad (3.62)$$

The epsilon symbol satisfies useful identities. One can show that the product of two epsilons with one index contracted is a sum of products of Kronecker deltas:

$$\epsilon_{ijk} \epsilon_{ipq} = \delta_{jp} \delta_{kq} - \delta_{jq} \delta_{kp}. \quad (3.63)$$

Its contraction (setting $p = j$) is also useful:

$$\epsilon_{ijk} \epsilon_{ijq} = 2\delta_{kq}. \quad (3.64)$$

The first of these two allows one to prove the familiar vector identity

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}. \quad (3.65)$$

It will be useful later on to consider the dot and cross products of *operator* triplets. Given the operators $\mathbf{X} = (\hat{X}_1, \hat{X}_2, \hat{X}_3)$ and $\mathbf{Y} = (\hat{Y}_1, \hat{Y}_2, \hat{Y}_3)$ we define

$$\begin{aligned} \mathbf{X} \cdot \mathbf{Y} &\equiv \hat{X}_i \hat{Y}_i, \\ (\mathbf{X} \times \mathbf{Y})_i &\equiv \epsilon_{ijk} \hat{X}_j \hat{Y}_k. \end{aligned} \quad (3.66)$$

In these definitions the order of the operators on the right hand side is as in the left-hand side. This is important to keep track of, since the \hat{X}_i and \hat{Y}_j operators may not commute. The dot product of two operator triplets is not necessarily commutative, nor is the cross product necessarily antisymmetric.

4 Spin states in arbitrary direction

We consider here the description and analysis of spin states that point in arbitrary directions, as specified by a unit vector \mathbf{n} :

$$\mathbf{n} = (n_x, n_y, n_z) = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta). \quad (4.67)$$

Here θ and ϕ are the familiar polar and azimuthal angles. We view the spatial vector \mathbf{n} as a triplet of numbers. Just like we did for $\boldsymbol{\sigma}$, we can define \mathbf{S} as the triplet of operators

$$\mathbf{S} = (\hat{S}_x, \hat{S}_y, \hat{S}_z). \quad (4.68)$$

Note that, in fact,

$$\mathbf{S} = \frac{\hbar}{2} \boldsymbol{\sigma}. \quad (4.69)$$

We can use \mathbf{S} to obtain, by a *dot* product with \mathbf{n} a spin operator $\hat{S}_{\mathbf{n}}$ that has a simple interpretation:

$$\hat{S}_{\mathbf{n}} \equiv \mathbf{n} \cdot \mathbf{S} \equiv n_x \hat{S}_x + n_y \hat{S}_y + n_z \hat{S}_z = \frac{\hbar}{2} \mathbf{n} \cdot \boldsymbol{\sigma}. \quad (4.70)$$

Note that $\hat{S}_{\mathbf{n}}$ is just an operator, or a hermitian matrix. We view $\hat{S}_{\mathbf{n}}$ as the spin operator in the direction of the unit vector \mathbf{n} . To convince you that this makes sense note that, for example, when \mathbf{n} points along z , we have $(n_x, n_y, n_z) = (0, 0, 1)$ and $\hat{S}_{\mathbf{n}}$ becomes \hat{S}_z . The same holds, of course, for the x and y directions. Moreover, just like all the \hat{S}_i , the eigenvalues of $\hat{S}_{\mathbf{n}}$ are $\pm \hbar/2$. This is needed physically, since all directions are physically equivalent and those two values for spin must be the only allowed values for all directions. To see that this is true we first compute the square of the matrix $\hat{S}_{\mathbf{n}}$:

$$(\hat{S}_{\mathbf{n}})^2 = \left(\frac{\hbar}{2}\right)^2 (\mathbf{n} \cdot \boldsymbol{\sigma})^2 = \left(\frac{\hbar}{2}\right)^2, \quad (4.71)$$

using (3.62). Moreover, since the Pauli matrices are traceless so is $\hat{S}_{\mathbf{n}}$:

$$\text{tr}(\hat{S}_{\mathbf{n}}) = n_i \text{tr}(\hat{S}_i) = n_i \frac{\hbar}{2} \text{tr}(\sigma_i) = 0. \quad (4.72)$$

By the same argument we used for Pauli matrices, we conclude that the eigenvalues of $\hat{S}_{\mathbf{n}}$ are indeed $\pm \hbar/2$. For an arbitrary direction we can write the matrix $\hat{S}_{\mathbf{n}}$ explicitly:

$$\begin{aligned} \hat{S}_{\mathbf{n}} &= \frac{\hbar}{2} \left[n_x \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + n_y \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + n_z \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right] \\ &= \frac{\hbar}{2} \begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix} \\ &= \frac{\hbar}{2} \begin{pmatrix} \cos \theta & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta \end{pmatrix}. \end{aligned} \quad (4.73)$$

Since the eigenvalues of $\hat{S}_{\mathbf{n}}$ are $\pm \hbar/2$ the associated spin eigenstates, denoted as $|\mathbf{n}; \pm\rangle$, satisfy

$$\hat{S}_{\mathbf{n}} |\mathbf{n}; \pm\rangle = \pm \frac{\hbar}{2} |\mathbf{n}; \pm\rangle. \quad (4.74)$$

The states $|\mathbf{n};+\rangle$ and $|\mathbf{n};-\rangle$ represent, respectively, a spin state that points up along \mathbf{n} , and a spin state that points down along \mathbf{n} . We can also find the eigenvalues of the matrix $\hat{S}_\mathbf{n}$ by direct computation. The eigenvalues are the roots of the equation $\det(\hat{S}_\mathbf{n} - \lambda \mathbf{1}) = 0$:

$$\det \begin{pmatrix} \frac{\hbar}{2} \cos \theta - \lambda & \frac{\hbar}{2} \sin \theta e^{-i\phi} \\ \frac{\hbar}{2} \sin \theta e^{i\phi} & -\frac{\hbar}{2} \cos \theta - \lambda \end{pmatrix} = \lambda^2 - \frac{\hbar^2}{4} (\cos^2 \theta + \sin^2 \theta) = \lambda^2 - \frac{\hbar^2}{4} = 0. \quad (4.75)$$

The eigenvalues are thus $\lambda = \pm \hbar/2$, as claimed. To find the eigenvector v associated with the eigenvalue λ we must solve the linear equation $(\hat{S}_\mathbf{n} - \lambda \mathbf{1})v = 0$. We denote by $|\mathbf{n};+\rangle$ the eigenvector associated with the eigenvalue $\hbar/2$. For this eigenvector we write the ansatz

$$|\mathbf{n};+\rangle = c_1 |+\rangle + c_2 |-\rangle = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad (4.76)$$

where for notational simplicity $|\pm\rangle$ refer to the states $|z;\pm\rangle$. The eigenvector equation becomes $(\hat{S}_\mathbf{n} - \frac{\hbar}{2}\mathbf{1})|\mathbf{n};+\rangle = 0$ and explicitly reads

$$\frac{\hbar}{2} \begin{pmatrix} \cos \theta - 1 & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta - 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0. \quad (4.77)$$

Either equation gives the same relation between c_1 and c_2 . The top equation, for example gives

$$c_2 = e^{i\phi} \frac{1 - \cos \theta}{\sin \theta} c_1 = e^{i\phi} \frac{\sin \frac{\theta}{2}}{\cos \frac{\theta}{2}} c_1. \quad (4.78)$$

(Check that the second equation gives the same relation.) We want normalized states, and therefore

$$|c_1|^2 + |c_2|^2 = 1 \rightarrow |c_1|^2 \left[1 + \frac{\sin^2 \frac{\theta}{2}}{\cos^2 \frac{\theta}{2}} \right] = 1 \rightarrow |c_1|^2 = \cos^2 \frac{\theta}{2}. \quad (4.79)$$

Since the overall phase of the eigenstate is not observable we take the simplest option for c_1 :

$$c_1 = \cos \frac{\theta}{2}, \quad c_2 = \sin \frac{\theta}{2} \exp(i\phi), \quad (4.80)$$

that is

$$|\mathbf{n};+\rangle = \cos \frac{\theta}{2} |+\rangle + \sin \frac{\theta}{2} e^{i\phi} |-\rangle. \quad (4.81)$$

As a quick check we see that for $\theta = 0$, which corresponds to a unit vector $\mathbf{n} = \mathbf{e}_3$ along the plus z direction we get $|\mathbf{e}_3;+\rangle = |+\rangle$. Note that even though ϕ is ambiguous when $\theta = 0$, this does not affect our answer, since the term with ϕ dependence vanishes. In the same way one can obtain the normalized eigenstate corresponding to $-\hbar/2$. A simple phase choice gives

$$|\mathbf{n};-\rangle = \sin \frac{\theta}{2} |+\rangle - \cos \frac{\theta}{2} e^{i\phi} |-\rangle. \quad (4.82)$$

If we again consider the $\theta = 0$ direction, this time the ambiguity of ϕ remains in the term that contains the $|z;-\rangle$ state. It is convenient to multiply this state by the phase $-e^{-i\phi}$. Doing this, the pair of

eigenstates read³

$ \mathbf{n}; +\rangle = \cos \frac{\theta}{2} +\rangle + \sin \frac{\theta}{2} e^{i\phi} -\rangle,$	(4.83)
$ \mathbf{n}; -\rangle = -\sin \frac{\theta}{2} e^{-i\phi} +\rangle + \cos \frac{\theta}{2} -\rangle.$	

The vectors are normalized. Furthermore, they are orthogonal

$$\langle \mathbf{n}; - | \mathbf{n}; + \rangle = -\sin \frac{\theta}{2} e^{i\phi} \cos \frac{\theta}{2} + \cos \frac{\theta}{2} \sin \frac{\theta}{2} e^{i\phi} = 0. \quad (4.84)$$

Therefore, $|\mathbf{n}; +\rangle$ and $|\mathbf{n}; -\rangle$ are an orthonormal pair of states.

Let us verify that the $|\mathbf{n}; \pm\rangle$ reduce to the known results as \mathbf{n} points along the z , x , and y axes. Again, if $\mathbf{n} = (0, 0, 1) = \mathbf{e}_3$, we have $\theta = 0$, and hence

$$|\mathbf{e}_3; +\rangle = |+\rangle, \quad |\mathbf{e}_3; -\rangle = |-\rangle, \quad (4.85)$$

which are, as expected, the familiar eigenstates of \hat{S}_z . If we point along the x axis, $\mathbf{n} = (1, 0, 0) = \mathbf{e}_1$ which corresponds to $\theta = \pi/2$, $\phi = 0$. Hence

$$|\mathbf{e}_1; +\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle) = |x; +\rangle, \quad |\mathbf{e}_1; -\rangle = \frac{1}{\sqrt{2}}(-|+\rangle + |-\rangle) = -|x; -\rangle, \quad (4.86)$$

where we compared with (2.32). Note that the second state came out with an overall minus sign. Since overall phases (or signs) are physically irrelevant, this is the expected answer: we got the eigenvectors of \hat{S}_x . Finally, if $\mathbf{n} = (0, 1, 0) = \mathbf{e}_2$, we have $\theta = \pi/2$, $\phi = \pi/2$ and hence, with $e^{\pm i\phi} = \pm i$, we have

$$|\mathbf{e}_2; +\rangle = \frac{1}{\sqrt{2}}(|+\rangle + i|-\rangle) = |y; +\rangle, \quad |\mathbf{e}_2; -\rangle = \frac{1}{\sqrt{2}}(i|+\rangle + |-\rangle) = i\frac{1}{\sqrt{2}}(|+\rangle - i|-\rangle) = i|y; -\rangle \quad (4.87)$$

which are, up to a phase for the second one, the eigenvectors of \hat{S}_y .

³The formula (4.83) works nicely at the north pole ($\theta = 0$), but at the south pole ($\theta = \pi$) the ϕ ambiguity shows up again. If one works near the south pole multiplying the results in (4.83) by suitable phases will do the job. The fact that no formula works well unambiguously through the full the sphere is not an accident.

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Chapter 2: Experiments with photons

B. Zwiebach
February 9, 2016

Contents

1 Mach-Zehnder Interferometer	1
2 Elitzur-Vaidman Bombs	6

1 Mach-Zehnder Interferometer

We have discussed before the Mach-Zehnder interferometer, which we show again in Figure 1. It contains two beam-splitters BS1 and BS2 and two mirrors. Inside the interferometer we have two beams, one going over the upper branch and one going over the lower branch. This extends beyond BS2: the upper branch continues to D0 while the lower branch continues to D1.

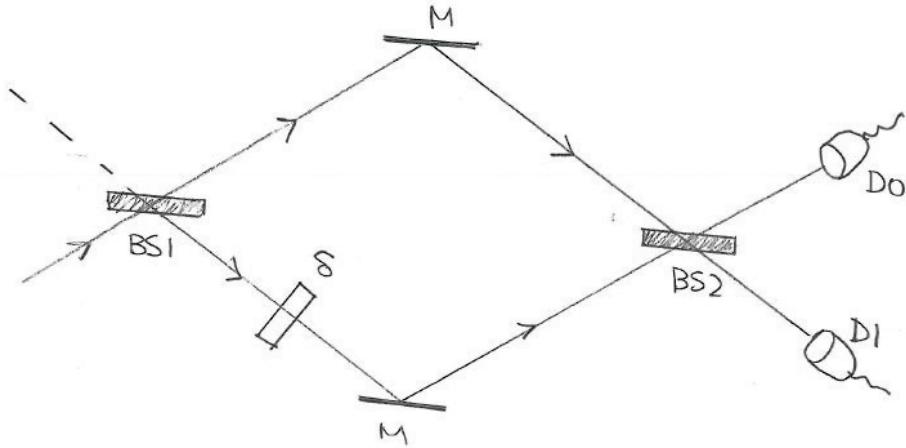


Figure 1: The Mach-Zehnder Interferometer

Vertical cuts in the above figure intersect the two beams and we can ask what is the probability to find a photon in each of the two beams at that cut. For this we need two probability *amplitudes*, or two complex numbers, whose norm-squared would give probabilities. We can encode this information in a two component vector as

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (1.1)$$

Here α is the probability amplitude to be in the upper beam and β the probability amplitude to be in the lower beam. Therefore, $|\alpha|^2$ would be the probability to find the photon in the upper beam and $|\beta|^2$ the probability to find the photon in the lower beam. Since the photon must be found in either one of the beams we must have

$$|\alpha|^2 + |\beta|^2 = 1. \quad (1.2)$$

Following this notation, we would have for the cases when the photon is definitely in one or the other beam:

$$\text{photon on upper beam: } \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{photon on bottom beam: } \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (1.3)$$

We can view the state (1.1) as a superposition of these two simpler states using the rules of vector addition and multiplication:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \beta \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (1.4)$$

In the interferometer shown in Figure 1 we included in the lower branch a ‘phase shifter’, a piece of



Figure 2: A phase shifter of phase factor $e^{i\delta}$. The amplitude gets multiplied by the phase.

material whose only effect is to multiply the probability amplitude by a fixed phase $e^{i\delta}$ with $\delta \in \mathbb{R}$. As shown in Figure 2, the probability amplitude α to the left of the device becomes $e^{i\delta}\alpha$ to the right of the device. Since the norm of a phase is one, the phase-shifter does not change the probability to find the photon. When the phase δ is equal to π the effect of the phase shifter is to change the sign of the wavefunction since $e^{i\pi} = -1$.

Let us now consider the effect of beam splitters in detail. If the incoming photon hits a beam-splitter from the top, we consider this photon to belong to the upper branch and represent it by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. If the incoming photon hits the beam-splitter from the bottom, we consider this photon to belong to the lower branch, and represent it by $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. We show the two cases in Figure 3. The effect of the beam splitter is to give an output wavefunction for each of the two cases:

$$\text{Left BS: } \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} s \\ t \end{pmatrix}, \quad \text{Right BS: } \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} u \\ v \end{pmatrix}. \quad (1.5)$$

As you can see from the diagram, for the photon hitting from above, s may be thought as a reflection amplitude and t as a transmission coefficient. Similarly, for the photon hitting from below, v may be thought as a reflection amplitude and u as a transmission coefficient. The four numbers s, t, u, v , by linearity, characterize completely the beam splitter. They can be used to predict the output given any incident photon, which may have amplitudes to hit both from above and from below. Indeed, an incident photon state $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ would give

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \alpha \begin{pmatrix} s \\ t \end{pmatrix} + \beta \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \alpha s + \beta u \\ \alpha t + \beta v \end{pmatrix} = \begin{pmatrix} s & u \\ t & v \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (1.6)$$

In summary, we see that the BS produces the following effect

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \rightarrow \begin{pmatrix} s & u \\ t & v \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (1.7)$$

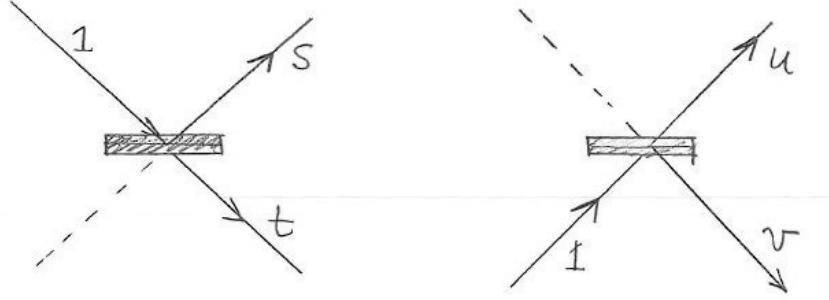


Figure 3: Left: A photon incident from the top; s and t are the reflected and transmitted amplitudes, respectively. Right: A photon incident from the bottom; v and u are the reflected and transmitted amplitudes, respectively.

We can represent the action of the beam splitter as matrix multiplication on the incoming wavefunction, with the two-by-two matrix

$$\begin{pmatrix} s & u \\ t & v \end{pmatrix}. \quad (1.8)$$

We must now figure out the constraints on s, t, u, v . Because probabilities must add up to one, equation (1.5) implies that

$$|s|^2 + |t|^2 = 1, \quad (1.9)$$

$$|u|^2 + |v|^2 = 1. \quad (1.10)$$

The kind of beam splitters we use are called balanced, which means that reflection and transmission probabilities are the same. So all four constants must have equal norm-squared:

$$|s|^2 = |t|^2 = |u|^2 = |v|^2 = \frac{1}{2}. \quad (1.11)$$

Let's try a guess for the values. Could we have

$$\begin{pmatrix} s & u \\ t & v \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} ? \quad (1.12)$$

This fails if acting on normalized wavefunctions (or column vectors) does not yield normalized wavefunctions. So we try with a couple of wavefunctions

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (1.13)$$

While the first example works out, the second does not, as $|1|^2 + |1|^2 = 2 \neq 1$. An easy fix is achieved by changing the sign of v :

$$\begin{pmatrix} s & u \\ t & v \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \quad (1.14)$$

Let's check that this matrix works in general. Thus acting on a state $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ with $|\alpha|^2 + |\beta|^2 = 1$ we find

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \alpha + \beta \\ \alpha - \beta \end{pmatrix}. \quad (1.15)$$

Indeed the resulting state is well normalized. The total probability is what we expect

$$\begin{aligned} \frac{1}{2}|\alpha + \beta|^2 + \frac{1}{2}|\alpha - \beta|^2 &= \frac{1}{2}(|\alpha|^2 + |\beta|^2 + \alpha\beta^* + \alpha^*\beta) + \frac{1}{2}(|\alpha|^2 + |\beta|^2 - \alpha\beta^* - \alpha^*\beta) \\ &= |\alpha|^2 + |\beta|^2 = 1. \end{aligned} \quad (1.16)$$

The minus sign in the bottom right entry of (1.14) means that a photon incident from below, as it is reflected, will have its amplitude changed by a sign or equivalently a phase shift by π (check this!). This effect, of course, is realized in practice. A typical beam splitter consists of a glass plate with a reflective dielectric coating on one side. The refractive index of the coating is chosen to be intermediate between that of glass and that of air. A reflection causes a phase shift only when light encounters a material of higher refractive index. This is the case in the transition of air to coating, but not in the transition from glass to coating. Thus the beam splitter represented by (1.14) would have its coating on the bottom side. Transmitted waves have no phase shift.

Another possibility for a beam splitter matrix is

$$\frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}, \quad (1.17)$$

which would be realized by a dielectric coating on the top side. You can quickly check that, like the previous matrix, its action also conserves probability. We will call the left beam-splitter BS1 and the right beam splitter BS2 and their respective matrices will be

$$\text{BS1 : } \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \text{BS2 : } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \quad (1.18)$$

The two beam splitters are combined to form the interferometer shown in Figure 4. If we now assume an input photon wavefunction $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ from the left, the output wavefunction that goes into the detectors is obtained by acting first with the BS1 matrix and then with the BS2 matrix:

$$\text{input : } \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \text{output: } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ -\alpha \end{pmatrix}. \quad (1.19)$$

With the help of this result, for any input photon state we can write immediately the output photon state that goes into the detectors.

If the input photon beam is $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, the output from the interferometer is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and therefore a photon will be detected at D0. This is shown in Figure 5. We can make a very simple table with the possible outcomes and their respective probabilities P :

Outcome	P
photon at D0	1
photon at D1	0

(1.20)

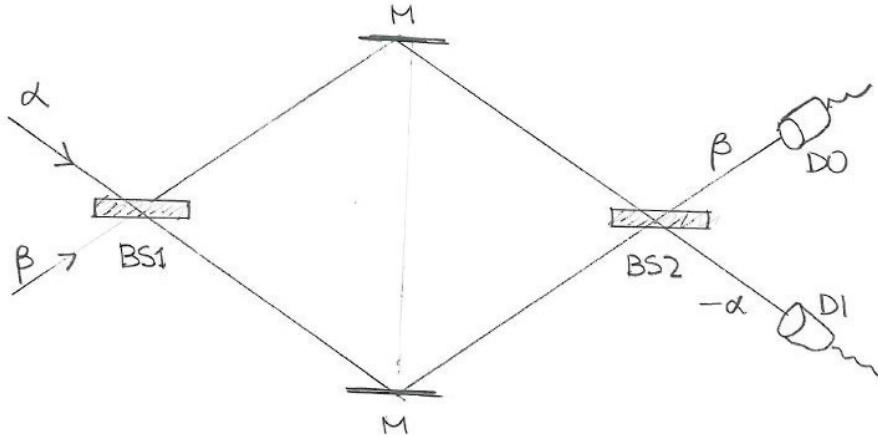


Figure 4: The Mach-Zehnder interferometer with input and output wavefunctions indicated.

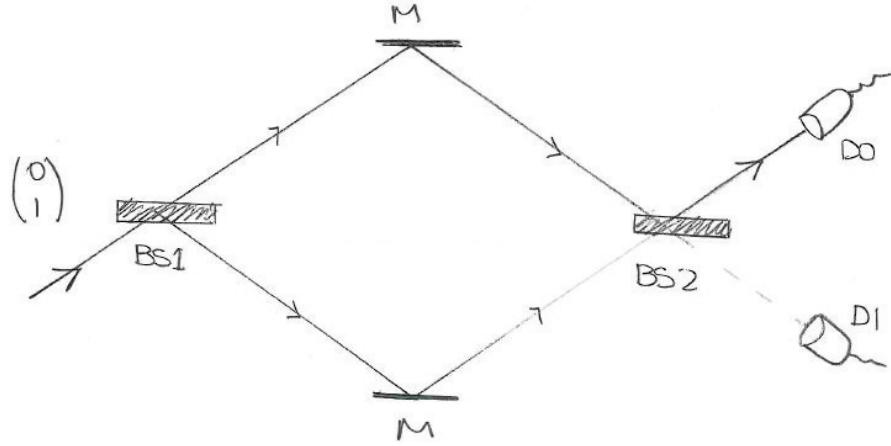


Figure 5: Incident photon from below will go into D0.

Now, block the lower path, as indicated in Figure 6. What happens then? It is best to track down things systematically. The input beam, acted by BS1 gives

$$\frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (1.21)$$

This is indicated in the figure, to the right of BS1. Then the lower branch is stopped, while the upper branch continues. The upper branch reaches BS2, and here the input is $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$, because nothing is coming from the lower branch. We therefore get an output

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}. \quad (1.22)$$

In this experiment there are three possible outcomes: the photon can be absorbed by the block, or

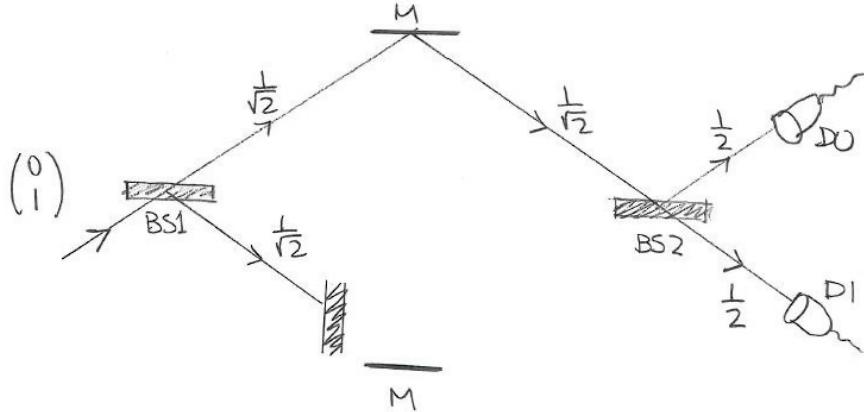


Figure 6: The probability to detect the photon at D1 can be changed by blocking one of the paths.

can go into any of the two detectors. As we see in the diagram, the probabilities are:

Outcome	P	
photon at block	$\frac{1}{2}$	
photon at D0	$\frac{1}{4}$	
photon at D1	$\frac{1}{4}$	

(1.23)

It is noteworthy that before blocking the lower path we could not get a photon to D1. The probability to reach D1 is now 1/4 and was increased by blocking a path.

2 Elitzur-Vaidman Bombs

To see that allowing the photon to reach D1 by blocking a path is very strange, we consider an imaginary situation proposed by physicists Avshalom Elitzur and Lev Vaidman, from Tel-Aviv University, in Israel. They imagined bombs with a special type of trigger: a photon detector. A narrow tube goes across each bomb and in the middle of the tube there is a photon detector. To detonate the bomb one sends a photon into the tube. The photon is then detected by the photon detector and the bomb explodes. If the photon detector is defective, however, the photon is not detected at all. It propagates freely through the tube and comes out of the bomb. The bomb does not explode.

Here is the situation we want to address. Suppose we have a number of Elitzur-Vaidman (EV) bombs, but we know that some of them have become defective. How could we tell if a bomb is operational without detonating it? Assume, for the sake of the problem, that we are unable to examine the detector without destroying the bomb.

We seem to be facing an impossible situation. If we send a photon into the detector tube and nothing happens we know the bomb is defective, but if the bomb is operational it would simply explode. It seems impossible to confirm that the photon detector in the bomb is working without testing it. Indeed, it is impossible in classical physics. It is not impossible in quantum mechanics, however. As we will see, we can perform what can be called an interaction-free measurement!

We now place an EV bomb on the lower path of the interferometer, with the detector tube properly aligned. Suppose we send in a photon as pictured. If the bomb is defective it is as if there is no detector, the lower branch of the interferometer is free and all the photons that we send in will end up in D0,

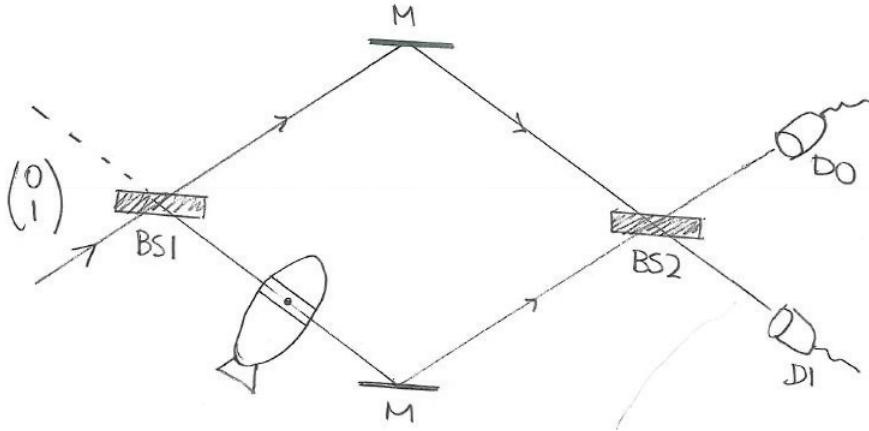


Figure 7: A Mach-Zehnder interferometer and an Elitzur-Vaidman bomb inserted on the lower branch, with the detector tube properly aligned. If the bomb is faulty all incident photons will end up at D0. If a photon ends up at D1 we know that the bomb is operational, even though the photon never went into the bomb detector!

just as they did in Figure 5.

Outcome	P	
photon at D0 no explosion	1	
photon at D1 no explosion	0	
bomb explodes	0	(2.24)

If the bomb is working, on the other hand, we have the situation we had in Figure 6, where we placed a block in the lower branch of the interferometer:

Outcome	P	
bomb explodes	$\frac{1}{2}$	
photon at D0 no explosion	$\frac{1}{4}$	
photon at D1 no explosion	$\frac{1}{4}$	(2.25)

Assume the bomb is working. Then 50% of the times the photon will hit it and it will explode, 25% of the time the photon will end in D0 and we can't tell if it is defective or not. But 25% of the time the photon will end in D1, and since this was impossible for a faulty bomb, we have learned that the bomb is operational! We have learned that even though the photon never made it through the bomb; it ended on D1. If you think about this you will surely realize it is extremely surprising and counterintuitive. But it is true, and experiments (without using bombs!) have confirmed that this kind of interaction-free measurement is indeed possible.

Sarah Geller transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare

<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

Lecture 6

B. Zwiebach
February 23, 2016

Contents

1 Normalization and time evolution	1
2 The Wavefunction as a Probability Amplitude	3
3 The Probability Current	5
4 Probability current in 3D and current conservation	7

1 Normalization and time evolution

The wavefunction $\Psi(x, t)$ that describes the quantum mechanics of a particle of mass m moving in a potential $V(x, t)$ satisfies the Schrödinger equation

$$i\hbar \frac{\partial \Psi(x, t)}{\partial t} = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x, t) \right) \Psi(x, t), \quad (1.1)$$

or more briefly

$$i\hbar \frac{\partial \Psi(x, t)}{\partial t} = \hat{H} \Psi(x, t). \quad (1.2)$$

The interpretation of the wavefunction arises by declaring that dP , defined by

$$dP = |\Psi(x, t)|^2 dx, \quad (1.3)$$

is the probability to find the particle in the interval dx centered on x at time t . It follows that the probabilities of finding the particle at all possible points must add up to one:

$$\int_{-\infty}^{\infty} \Psi^*(x, t) \Psi(x, t) dx = 1. \quad (1.4)$$

We will try to understand how this equation is compatible with the time evolution prescribed by the Schrödinger equation. But before that let us examine what kind of conditions are required from wavefunctions in order to satisfy (1.4).

Suppose the wavefunction has well-defined limits as $x \rightarrow \pm\infty$. If those limits are different from zero, the integral around infinity would produce an infinite result, which is inconsistent with the claim that the total integral is one. Therefore the limits should be zero:

$$\boxed{\lim_{x \rightarrow \pm\infty} \Psi(x, t) = 0.} \quad (1.5)$$

It is in principle possible to have a wavefunction that has no well-defined limit at infinity but is still square integrable. But such cases do not seem to appear in practice so we will assume that (1.5) holds. It would also be natural to assume that the spatial derivative of Ψ vanishes as $x \rightarrow \pm\infty$ but, as we will see soon, it suffices to assume that the limit of the spatial derivative of Ψ is bounded

$$\boxed{\lim_{x \rightarrow \pm\infty} \frac{\partial\Psi(x, t)}{\partial x} < \infty.} \quad (1.6)$$

We have emphasized before that the overall numerical factor multiplying the wavefunction is not physical. But equation (1.4) seems to be in conflict with this: if a given Ψ satisfies it, the presumed equivalent 2Ψ will not! To make precise sense of probabilities it is *convenient* to work with normalized wavefunctions, but it is not necessary, as we show now. Since time plays no role in the argument, so assume in all that follows that the equations refer to some time t_0 arbitrary but fixed. Suppose you have a wavefunction such that

$$\int dx |\Psi|^2 = \mathcal{N} \neq 1. \quad (1.7)$$

Then I claim that the probability dP to find the particle in the interval dx about x is given by

$$dP = \frac{1}{\mathcal{N}} |\Psi|^2 dx. \quad (1.8)$$

This is consistent because

$$\int dP = \frac{1}{\mathcal{N}} \int dx |\Psi|^2 = \frac{1}{\mathcal{N}} \cdot \mathcal{N} = 1. \quad (1.9)$$

Note that dP is not changed when Ψ is multiplied by any number. Thus, this picture makes it clear that the overall scale of Ψ contains no physics. As long as the integral $\int |\Psi|^2 dx < \infty$ the wavefunction is said to be **normalizable, or square-integrable**. By adjusting the overall coefficient of Ψ we can then make it **normalized**. Indeed, again assuming (1.7) the new wavefunction Ψ' defined by

$$\Psi' = \frac{1}{\sqrt{\mathcal{N}}} \Psi, \quad (1.10)$$

is properly normalized. Indeed

$$\int dx |\Psi'|^2 = \frac{1}{\mathcal{N}} \int |\Psi|^2 dx = 1. \quad (1.11)$$

We sometimes work with wavefunctions for which the integral (1.4) is infinite. Such wavefunctions can be very useful. In fact, the de Broglie plane wave $\Psi = \exp(ikx - i\omega t)$ for a free particle is a good example: since $|\Psi|^2 = 1$ the integral is in fact infinite. What this means is that $\exp(ikx - i\omega t)$ does not truly represent a single particle. To construct a square-integrable wavefunction we can use a superposition of plane waves. It is indeed a pleasant surprise that the superposition of infinitely many non-square integrable waves is square integrable!

2 The Wavefunction as a Probability Amplitude

Let's begin with a normalized wavefunction at initial time t_0

$$\int_{-\infty}^{\infty} \Psi^*(x, t_0) \Psi(x, t_0) dx = 1. \quad (2.1)$$

Since $\Psi(x, t_0)$ and the Schrödinger equation determine Ψ for all times, do we then have

$$\int_{-\infty}^{\infty} \Psi^*(x, t) \Psi(x, t) dx = 1 ? \quad (2.2)$$

Define the **probability density** $\rho(x, t)$

$$\rho(x, t) \equiv \Psi^*(x, t) \Psi(x, t) = |\Psi(x, t)|^2. \quad (2.3)$$

Define also $\mathcal{N}(t)$ as the integral of the probability density throughout space:

$$\mathcal{N}(t) \equiv \int \rho(x, t) dx. \quad (2.4)$$

The statement in (2.1) that the wavefunction begins well normalized is

$$\mathcal{N}(t_0) = 1, \quad (2.5)$$

and the condition that it remain normalized for all later times is $\mathcal{N}(t) = 1$. This would be guaranteed if we showed that for all times

$$\frac{d\mathcal{N}(t)}{dt} = 0. \quad (2.6)$$

We call this *conservation* of probability. Let's check if the Schrödinger equation ensures this condition will hold:

$$\begin{aligned} \frac{d\mathcal{N}(t)}{dt} &= \int_{-\infty}^{\infty} \frac{\partial \rho(x, t)}{\partial t} dx \\ &= \int_{-\infty}^{\infty} \left(\frac{\partial \Psi^*}{\partial t} \Psi(x, t) + \Psi^*(x, t) \frac{\partial \Psi(x, t)}{\partial t} \right) dx. \end{aligned} \quad (2.7)$$

From the Schrödinger equation, and its complex conjugate

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi \implies \frac{\partial \Psi}{\partial t} = -\frac{i}{\hbar} \hat{H}\Psi, \quad (2.8)$$

$$-i\hbar \frac{\partial \Psi^*}{\partial t} = (\hat{H}\Psi)^* \implies \frac{\partial \Psi^*}{\partial t} = \frac{i}{\hbar} (\hat{H}\Psi)^*. \quad (2.9)$$

In complex conjugating the Schrödinger equation we used that the complex conjugate of the time derivative of Ψ is simply the time derivative of the complex conjugate of Ψ . To conjugate

the right hand side we simply added the star to the whole of $\hat{H}\Psi$. We now use (2.8) and (2.9) in (2.7) to find

$$\begin{aligned}\frac{d\mathcal{N}(t)}{dt} &= \int_{-\infty}^{\infty} \left(\frac{i}{\hbar}(\hat{H}\Psi)^* \Psi - \frac{i}{\hbar}\Psi^*(\hat{H}\Psi) \right) dx \\ &= \frac{i}{\hbar} \left(\int_{-\infty}^{\infty} (\hat{H}\Psi)^* \Psi dx - \int_{-\infty}^{\infty} \Psi^*(\hat{H}\Psi) dx \right).\end{aligned}\quad (2.10)$$

To show that the time derivative of $\mathcal{N}(t)$ vanishes, it suffices to show that

$$\boxed{\int_{-\infty}^{\infty} (\hat{H}\Psi)^* \Psi = \int_{-\infty}^{\infty} \Psi^*(\hat{H}\Psi).}\quad (2.11)$$

Equation (2.11) is the condition on the Hamiltonian operator \hat{H} for conservation of probability. In fact, if \hat{H} is a Hermitian operator the condition will be satisfied. The operator \hat{H} is a **Hermitian** operator if it satisfies

$$\boxed{\text{Hermitian operator: } \int_{-\infty}^{\infty} (\hat{H}\Psi_1)^* \Psi_2 = \int_{-\infty}^{\infty} \Psi_1^*(\hat{H}\Psi_2).}\quad (2.12)$$

Here we have two wavefunctions that are arbitrary, but satisfy the conditions (1.5) and (1.6). As you can see, a Hermitian operator can be switched from acting on the first function to acting on the second function. When the two functions are the same, we recover condition (2.11).

It is worth closing this circle of ideas by defining the **Hermitian conjugate** T^\dagger of the linear operator T . This is done as follows:

$$\int_{-\infty}^{\infty} \Psi_1^*(T\Psi_2) = \int_{-\infty}^{\infty} (T^\dagger \Psi_1)^* \Psi_2.\quad (2.13)$$

The operator T^\dagger , which is also linear, is calculated by starting from the left-hand side and trying to recast the expression with no operator acting on the second function. An operator T is said to be Hermitian if it is equal to its Hermitian conjugate:

$$T \text{ is Hermitian if } T^\dagger = T.\quad (2.14)$$

Hermitian operators are very important in Quantum Mechanics. They have real eigenvalues and one can always find a basis of the state space in terms of orthonormal eigenstates. It turns out that observables in Quantum Mechanics are represented by Hermitian operators, and the possible measured values of those observables are given by their eigenvalues. Our quest to show that normalization is preserved under time evolution in Quantum Mechanics has come down to showing that the Hamiltonian operator is Hermitian.

3 The Probability Current

Let's take a closer look at the integrand of equation (2.10). Using the explicit expression for the Hamiltonian we have

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= \frac{i}{\hbar}((\hat{H}\Psi)^* \Psi - \Psi^*(\hat{H}\Psi)) \\ &= \frac{i}{\hbar} \left[-\frac{\hbar^2}{2m} \left(\frac{\partial^2 \Psi^*}{\partial x^2} \Psi - \Psi^* \frac{\partial^2 \Psi}{\partial x^2} \right) + V(x, t) \Psi^* \Psi - \Psi^* V(x, t) \Psi \right].\end{aligned}\quad (3.1)$$

The contributions from the potential cancel and we then get

$$\frac{i}{\hbar}((\hat{H}\Psi)^* \Psi - \Psi^*(\hat{H}\Psi)) = \frac{\hbar}{2im} \left(\frac{\partial^2 \Psi^*}{\partial x^2} \Psi - \Psi^* \frac{\partial^2 \Psi}{\partial x^2} \right). \quad (3.2)$$

The only chance to get to show that the integral of the right-hand side is zero is to show that it is a total derivative. Indeed, it is!

$$\begin{aligned}\frac{i}{\hbar}((\hat{H}\Psi)^* \Psi - \Psi^*(\hat{H}\Psi)) &= \frac{\partial}{\partial x} \left[\frac{\hbar}{2im} \left(\frac{\partial \Psi^*}{\partial x} \Psi - \Psi^* \frac{\partial \Psi}{\partial x} \right) \right] \\ &= -\frac{\partial}{\partial x} \left[\frac{\hbar}{2im} \left(\Psi^* \frac{\partial \Psi}{\partial x} - \frac{\partial \Psi^*}{\partial x} \Psi \right) \right] \\ &= -\frac{\partial}{\partial x} \left[\frac{\hbar}{2im} 2i \operatorname{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right) \right] \\ &= -\frac{\partial}{\partial x} \left[\frac{\hbar}{m} \operatorname{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right) \right],\end{aligned}\quad (3.3)$$

where we used that $z - z^* = 2i \operatorname{Im}(z)$. Recall that the left-hand side we have evaluated is actually $\frac{\partial \rho}{\partial t}$ and therefore the result obtained so far is

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} \left[\frac{\hbar}{m} \operatorname{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right) \right] = 0. \quad (3.4)$$

This equation encodes charge conservation and is of the type

$$\frac{\partial \rho}{\partial t} + \frac{\partial J}{\partial x} = 0, \quad (3.5)$$

where $J(x, t)$ is the current associated with the charge density ρ . We have therefore identified a probability current

$$J(x, t) \equiv \frac{\hbar}{m} \operatorname{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right).$$

(3.6)

There is just one component for this current since the particle moves in one dimension. The units of J are one over time, or probability per unit time, as we now verify.

For one spatial dimension, $[\Psi] = L^{-1/2}$, which is easily seen from the requirement that $\int dx |\Psi|^2$ is unit free. (When working with d spatial dimensions the wavefunction will have units of $L^{-d/2}$). We then have

$$\left[\Psi^* \frac{\partial \Psi}{\partial x} \right] = \frac{1}{L^2}, \quad [\hbar] = \frac{ML^2}{T}, \quad \left[\frac{\hbar}{m} \right] = \frac{L^2}{T}, \quad (3.7)$$

$$\implies [J] = \frac{1}{T} = \text{probability per unit time} \quad (3.8)$$

We can now show that the time derivative of \mathcal{N} is zero. Indeed, using (3.5) we have

$$\frac{d\mathcal{N}}{dt} = \int_{-\infty}^{\infty} dx \frac{\partial \rho}{\partial t} = - \int_{\infty}^{\infty} \frac{\partial J}{\partial x} dx = -(J(\infty, t) - J(-\infty, t)). \quad (3.9)$$

The derivative vanishes if the probability current vanishes at infinity. Recalling that

$$J = \frac{\hbar}{2im} \left(\Psi^* \frac{\partial \Psi}{\partial x} - \Psi \frac{\partial \Psi^*}{\partial x} \right), \quad (3.10)$$

we see that the current indeed vanishes because we restrict ourselves to wavefunctions for which $\lim_{x \rightarrow \pm\infty} \Psi = 0$ and $\lim_{x \rightarrow \pm\infty} \frac{\partial \Psi}{\partial x}$ remains bounded. We therefore have

$$\frac{d\mathcal{N}}{dt} = 0, \quad (3.11)$$

as we wanted to show.

To illustrate how probability conservation works more generally in one dimension, focus on a segment $x \in [a, b]$. Then the probability P_{ab} to find the particle in the segment $[a, b]$, is given by

$$P_{ab} = \int_a^b \rho(x, t) dx. \quad (3.12)$$

If we now take the time derivative of this and, as before, use current conservation we get

$$\frac{dP_{ab}}{dt} = - \int_a^b \frac{\partial J(x, t)}{\partial x} dt = -J(b, t) + J(a, t). \quad (3.13)$$

This is the expected result. If the amount of probability in the region $[a, b]$ changes in time, it must be due to probability current flowing in or out at the edges of the interval. Assuming the currents at $x = b$ and at $x = a$ are positive, we note that probability is flowing out at $x = b$ and is coming in at $x = a$. The signs in the above right-hand side correctly reflect the effect of these flows on the rate of change of the total probability inside the segment.

4 Probability current in 3D and current conservation

The determination of the probability current \mathbf{J} for a particle moving in three dimensions follows the route taken before, but we use the 3D version of the Schrödinger equation. After some work (homework) the probability density and the current are determined to be

$$\rho(\mathbf{x}, t) = |\Psi(\mathbf{x}, t)|^2, \quad \mathbf{J}(x, t) = \frac{\hbar}{m} \operatorname{Im} (\Psi^* \nabla \Psi), \quad (4.1)$$

and satisfy the conservation equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0. \quad (4.2)$$

In three spatial dimensions, $[\Psi] = L^{-\frac{3}{2}}$ and the units of \mathbf{J} are quickly determined

$$[\Psi^* \nabla \Psi] = \frac{1}{L^4}, \quad \left[\frac{\hbar}{m} \right] = \frac{L^2}{T} \quad (4.3)$$

$$\implies [\mathbf{J}] = \frac{1}{TL^2} = \text{probability per unit time per unit area} \quad (4.4)$$

The conservation equation (4.2) is particularly clear in integral language. Consider a fixed region V of space and the probability $Q_V(t)$ to find the particle inside the region:

$$Q_V(t) = \int_V \rho(\mathbf{x}, t) d^3 \mathbf{x}. \quad (4.5)$$

The time derivative of the probability is then calculated using the conservation equation

$$\frac{dQ_V}{dt} = \int_V \frac{\partial \rho}{\partial t} d^3 \mathbf{x} = - \int_V \nabla \cdot \mathbf{J} d^3 \mathbf{x}. \quad (4.6)$$

Finally, using Gauss' law we find

$$\frac{dQ_V}{dt} = - \int_S \mathbf{J} \cdot d\mathbf{a}, \quad (4.7)$$

where S is the boundary of the volume V . The interpretation here is clear. The probability that the particle is inside V may change in time if there is flux of the probability current across the boundary of the region. When the volume extends throughout space, the boundary is at infinity, and the conditions on the wavefunction (which we have not discussed in the 3D case) imply that the flux across the boundary at infinity vanishes.

Our probability density, probability current, and current conservation are in perfect analogy to electromagnetic charge density, current density, and current conservation. In electromagnetism charges flow, in quantum mechanics probability flows. The terms of the correspondence are summarized by the table.

	Electromagnetism	Quantum Mechanics
ρ	charge density	probability density
Q_V	charge in a volume V	probability to find particle in V
\mathbf{J}	current density	probability current density

Sarah Geller transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare
<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

Lectures 21 and 22: Hydrogen Atom

B. Zwiebach
May 4, 2016

Contents

1 The Hydrogen Atom	1
2 Hydrogen atom spectrum	4

1 The Hydrogen Atom

Our goal here is to show that the two-body quantum mechanical problem of the hydrogen atom can be recast as one in which we have center-of-mass degrees of freedom that behave like a free particle and relative-motion degrees of freedom for which we have dynamics controlled by a central potential.

The hydrogen atom consists of a proton and an electron moving in three dimensions. We label the position and momentum operators of the proton as $\hat{\mathbf{x}}_p, \hat{\mathbf{p}}_p$, and those of the electron as $\hat{\mathbf{x}}_e, \hat{\mathbf{p}}_e$. These are canonical variables, meaning they satisfy the canonical commutation relations:

$$[(\hat{\mathbf{x}}_p)_i, (\hat{\mathbf{p}}_p)_j] = i\hbar\delta_{ij}, \quad [(\hat{\mathbf{x}}_e)_i, (\hat{\mathbf{p}}_e)_j] = i\hbar\delta_{ij}. \quad (1.1)$$

Here the subscripts $i, j = 1, 2, 3$ denote the various components of the vector operators. Furthermore, the proton variables **commute** with the electron variables. We have two pairs of *independent* canonical variables.

The wavefunction for the system is a function of the positions of both particles:

$$\Psi(\mathbf{x}_p, \mathbf{x}_e), \quad (1.2)$$

and the quantity

$$|\Psi(\mathbf{x}_p, \mathbf{x}_e)|^2 d^3\mathbf{x}_p d^3\mathbf{x}_e, \quad (1.3)$$

is the probability to find the proton within a window $d^3\mathbf{x}_p$ of \mathbf{x}_p and the electron within a window $d^3\mathbf{x}_e$ of \mathbf{x}_e . The Hamiltonian of the system is given by

$$\hat{H} = \frac{\hat{\mathbf{p}}_p^2}{2m_p} + \frac{\hat{\mathbf{p}}_e^2}{2m_e} + V(|\mathbf{x}_e - \mathbf{x}_p|). \quad (1.4)$$

Note that the kinetic energy is simply the sum of the kinetic energy of the proton and kinetic energy of the electron. The potential only depends on the magnitude of the separation between the two particles, not on their individual positions.

In order to simplify the problem, we will introduce two new pairs of independent canonical variables. The first pair is associated with the center-of-mass (CM) motion. We introduce the total momentum operator $\hat{\mathbf{P}}$ and the CM position operator $\hat{\mathbf{X}}$, given by

$$\hat{\mathbf{P}} = \hat{\mathbf{p}}_p + \hat{\mathbf{p}}_e, \quad \hat{\mathbf{X}} = \frac{m_e \hat{\mathbf{x}}_e + m_p \hat{\mathbf{x}}_p}{m_e + m_p}. \quad (1.5)$$

The operator $\hat{\mathbf{X}}$ is given by the typical expression for the center-of-mass of the system, but with the positions replaced by position operators. Using the commutation relations (1.1), we can show that $\hat{\mathbf{X}}$ and $\hat{\mathbf{P}}$ are canonical conjugates:

$$\begin{aligned} [(\hat{\mathbf{X}})_i, (\hat{\mathbf{P}})_j] &= \left[\frac{m_e(\hat{\mathbf{x}}_e)_i + m_p(\hat{\mathbf{x}}_p)_i}{m_e + m_p}, (\hat{\mathbf{p}}_p)_j + (\hat{\mathbf{p}}_e)_j \right] \\ &= \frac{m_e}{m_e + m_p} [(\hat{\mathbf{x}}_e)_i, (\hat{\mathbf{p}}_e)_j] + \frac{m_p}{m_e + m_p} [(\hat{\mathbf{x}}_p)_i, (\hat{\mathbf{p}}_p)_j] \\ &= \frac{m_e}{m_e + m_p} i\hbar\delta_{ij} + \frac{m_p}{m_e + m_p} i\hbar\delta_{ij}, \end{aligned} \quad (1.6)$$

resulting in the expected

$$[(\hat{\mathbf{X}})_i, (\hat{\mathbf{P}})_j] = i\hbar\delta_{ij}. \quad (1.7)$$

For the second pair of canonical variables we will define relative position and momentum operators. The relative position operator is the natural variable implied by the form of the potential:

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_e - \hat{\mathbf{x}}_p. \quad (1.8)$$

Since the second pair of canonical variables must commute with the first pair, we must check that \mathbf{x} defined above, commutes with \mathbf{X} and with \mathbf{P} . The commutation with \mathbf{X} is automatic and the commutation with \mathbf{P} works thanks to the minus sign in the above definition. We must now construct a relative momentum operator $\hat{\mathbf{p}}$ that is canonically conjugate to \mathbf{x} . It must be built from the momentum operators of the two particles, so we write

$$\hat{\mathbf{p}} = \alpha \hat{\mathbf{p}}_e - \beta \hat{\mathbf{p}}_p, \quad (1.9)$$

with α and β coefficients to be determined. To be canonically conjugate, the relative operators must satisfy

$$[(\hat{\mathbf{x}})_i, (\hat{\mathbf{p}})_j] = i\hbar\delta_{ij} \rightarrow \alpha + \beta = 1, \quad (1.10)$$

using the above definitions of $\hat{\mathbf{x}}$ and $\hat{\mathbf{p}}$ and the proton and electron commutators. Finally, the relative momentum must commute with the CM coordinate

$$[(\hat{\mathbf{X}})_i, (\hat{\mathbf{p}})] = 0 \rightarrow m_e\alpha - m_p\beta = 0. \quad (1.11)$$

The two equations for α and β can be solved to find

$$\alpha = \frac{m_p}{m_e + m_p}, \quad \beta = \frac{m_e}{m_e + m_p}. \quad (1.12)$$

We define the total mass M and the reduced mass μ as follows

$$M = m_e + m_p, \quad \mu = \frac{m_e m_p}{m_e + m_p}. \quad (1.13)$$

The reduced mass of a pair of particles with very different masses is approximately equal to the mass of the lower-mass particle. Using these definitions

$$\alpha = \frac{\mu}{m_e}, \quad \beta = \frac{\mu}{m_p}. \quad (1.14)$$

Thus, collecting the relative variables we have

$$\hat{\mathbf{p}} = \mu \left(\frac{\hat{\mathbf{p}}_e}{m_e} - \frac{\hat{\mathbf{p}}_p}{m_p} \right) = \frac{m_p}{M} \hat{\mathbf{p}}_e - \frac{m_e}{M} \hat{\mathbf{p}}_p, \quad \hat{\mathbf{x}} = \hat{\mathbf{x}}_e - \hat{\mathbf{x}}_p. \quad (1.15)$$

Note that the relative momentum \mathbf{p} can be written in terms of velocities as follows: $\mathbf{p} = \mu(\mathbf{v}_e - \mathbf{v}_p)$. The relative momentum vanishes if the motion is only CM motion, in which case the velocities of the two parties are the same.

We can now rewrite the Hamiltonian in terms of the new variables. Solving for the original momentum operators in terms of $\hat{\mathbf{P}}$ and $\hat{\mathbf{p}}$, we find

$$\hat{\mathbf{p}}_p = \frac{m_p}{M} \hat{\mathbf{P}} - \hat{\mathbf{p}}, \quad \hat{\mathbf{p}}_e = \frac{m_e}{M} \hat{\mathbf{P}} + \hat{\mathbf{p}}. \quad (1.16)$$

We can then rewrite the kinetic terms of the Hamiltonian in the form

$$\begin{aligned} \frac{\hat{\mathbf{p}}_p^2}{2m_p} + \frac{\hat{\mathbf{p}}_e^2}{2m_e} &= \frac{1}{2m_p} \left(\frac{m_p^2}{M^2} \hat{\mathbf{P}}^2 - \frac{2m_p}{M} \hat{\mathbf{P}} \cdot \hat{\mathbf{p}} + \hat{\mathbf{p}}^2 \right) \\ &\quad + \frac{1}{2m_e} \left(\frac{m_e^2}{M^2} \hat{\mathbf{P}}^2 + \frac{2m_e}{M} \hat{\mathbf{P}} \cdot \hat{\mathbf{p}} + \hat{\mathbf{p}}^2 \right) \\ &= \frac{\hat{\mathbf{P}}^2}{2M} + \frac{\hat{\mathbf{p}}^2}{2\mu}. \end{aligned} \quad (1.17)$$

Happily the term coupling the two momenta vanishes. Thus the center of mass degrees of freedom and the relative degrees of freedom give independent contributions to the kinetic energy. The Hamiltonian can then be written as

$$\hat{H} = \frac{\hat{\mathbf{P}}^2}{2M} + \frac{\hat{\mathbf{p}}^2}{2\mu} + V(\hat{\mathbf{x}}). \quad (1.18)$$

In position space, the total and relative momentum operators can be expressed as gradients

$$\hat{\mathbf{P}} \rightarrow \frac{\hbar}{i} \nabla_{\mathbf{X}}, \quad \hat{\mathbf{p}} \rightarrow \frac{\hbar}{i} \nabla_{\mathbf{x}}. \quad (1.19)$$

Each ∇ has a subscript indicating the type of coordinate we use to take the derivatives. Just like we had a wavefunction $\Psi(\mathbf{x}_e, \mathbf{x}_p)$ the new canonical variables require that we now think of the wavefunction as a function $\Psi(\mathbf{X}, \mathbf{x})$ of the new coordinates.

We solve the time-independent Schrödinger equation by using separation of variables

$$\Psi(\mathbf{X}, \mathbf{x}) = \Psi_{\text{CM}}(\mathbf{X}) \Psi_{\text{rel}}(\mathbf{x}). \quad (1.20)$$

Plugging these into the time-independent Schrödinger equation $\hat{H}\Psi = E\Psi$, we reach

$$\left[\frac{\hat{\mathbf{P}}^2}{2M} \Psi_{\text{CM}}(\mathbf{X}) \right] \Psi_{\text{rel}}(\mathbf{x}) + \left[\frac{\hat{\mathbf{p}}^2}{2\mu} \Psi_{\text{rel}}(\mathbf{x}) + V(|\hat{\mathbf{x}}|) \Psi_{\text{rel}}(\mathbf{x}) \right] \Psi_{\text{CM}}(\mathbf{X}) = E \Psi_{\text{CM}}(\mathbf{X}) \Psi_{\text{rel}}(\mathbf{x}). \quad (1.21)$$

Dividing by the total wavefunction $\Psi_{\text{CM}}(\mathbf{X}) \Psi_{\text{rel}}(\mathbf{x})$, this becomes

$$\frac{1}{\Psi_{\text{CM}}(\mathbf{X})} \left[\frac{\hat{\mathbf{P}}^2}{2M} \Psi_{\text{CM}}(\mathbf{X}) \right] + \frac{1}{\Psi_{\text{rel}}(\mathbf{x})} \left[\frac{\hat{\mathbf{p}}^2}{2\mu} + V(|\hat{\mathbf{x}}|) \right] \Psi_{\text{rel}}(\mathbf{x}) = E. \quad (1.22)$$

The first term on the left-hand side is a function of \mathbf{X} only and the second term on the left-hand side is a function of \mathbf{x} only. Their sum is equal to the constant E and since \mathbf{x} and \mathbf{X} are independent variables, each term must individually be constant. We thus set the first term equal to the constant E_{CM} and the second term equal to the constant E_{rel} , resulting in the following equations:

$$\frac{\hat{\mathbf{P}}^2}{2M} \Psi_{\text{CM}}(\mathbf{X}) = E_{\text{CM}} \Psi_{\text{CM}}(\mathbf{X}), \quad (1.23)$$

$$\left[\frac{\hat{\mathbf{p}}^2}{2\mu} + V(|\mathbf{x}|) \right] \Psi_{\text{rel}}(\mathbf{x}) = E_{\text{rel}} \Psi_{\text{rel}}(\mathbf{x}), \quad (1.24)$$

$$E = E_{\text{CM}} + E_{\text{rel}}. \quad (1.25)$$

We get two Schrödinger equations. The first equation tells us that the center of mass moves as a free particle of mass M . Thus, the CM energy is not quantized and we get plane wave solutions. The second equation is for the relative motion, and as we wanted to show, it is described as motion in a central potential. The third equation tells us that the total energy is the sum of the center-of-mass energy and the energy from the relative motion.

2 Hydrogen atom spectrum

We now have the tools to study the hydrogen atom, which has a central potential given by

$$V(r) = -\frac{Ze^2}{r}, \quad (2.1)$$

where Z is the number of protons in the nucleus. For hydrogen we have $Z = 1$. But it is worth considering $Z > 1$ in which case we are describing the motion of an electron around the nucleus of some other atom. We will also define following physical constants.

- The fine structure constant α : $\alpha = \frac{e^2}{\hbar c} \simeq \frac{1}{137}$.
- The *Bohr radius* a_0 . This is the characteristic length scale in the problem. It can be calculated by equating kinetic and potential energies expressed in terms of a_0 and ignoring all numerical constants:

$$\frac{\hbar^2}{m_e a_0^2} = \frac{e^2}{a_0}. \quad (2.2)$$

Here the mass should be the reduced mass, which in this case can be taken rather accurately to be the mass of the electron. We then have explicitly,

$$\begin{aligned} a_0 &= \frac{\hbar^2}{me^2} = \frac{\hbar^2 c^2}{e^2 mc^2} = \frac{\hbar c}{\left(\frac{e^2}{\hbar c}\right) mc^2} = \frac{\hbar c}{\alpha mc^2} \\ &= \frac{197 \text{ MeV fm}}{0.51 \times 10^6 \text{ eV} \left(\frac{1}{137}\right)} = \frac{1970 \text{ eV Angstrom}}{0.51 \times 10^6 \text{ eV}} \times 137 \\ &= 0.529 \text{ Angstroms} \simeq 53 \text{ pm}. \end{aligned} \quad (2.3)$$

For the energy scale estimate we have

$$\frac{e^2}{a_0} = e^2 \left(\frac{me^2}{\hbar^2} \right) = \left(\frac{e^4}{\hbar^2 c^2} \right) mc^2 = \alpha^2 mc^2 = \frac{1}{(137)^2} \times (511\,000 \text{ eV}) \simeq 27.2 \text{ eV} \quad (2.4)$$

There are other characteristic lengths that are interesting:

$$\begin{aligned}\alpha a_0 &= \text{Compton wavelength of electron} = \lambda_e \simeq 390 \text{ fm}, \\ \alpha^2 a_0 &= \text{classical electron radius} \simeq 2.8 \text{ fm}.\end{aligned}\quad (2.5)$$

Let us do some work now! The radial Schrödinger equation for bound states $E < 0$ reads

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dr^2} + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} - \frac{Ze^2}{r} \right) u = Eu. \quad (2.6)$$

We could label the wavefunction u as $u_{E\ell}$ as the solutions will certainly depend on ℓ and the energy E . As usual, we like to work with a unit free coordinate. This could be achieved by writing $r = a_0 x$, with x unit free and a_0 carrying the length units of r . It will be more convenient to use a slight variation to eliminate Z from the equation and some factors of two. We will take the new unit-free coordinate x to be defined from

$$r \equiv \frac{a_0}{2Z} x. \quad (2.7)$$

The Schrödinger equation then becomes

$$\begin{aligned}&\left(-\frac{\hbar^2}{2m} \frac{4Z^2}{a_0^2} \frac{d^2}{dx^2} + \frac{4Z^2}{a_0^2} \frac{\hbar^2}{2m} \frac{l(l+1)}{x^2} - \frac{2Z^2 e^2}{a_0} \frac{1}{x} \right) u = Eu \\ &\rightarrow \left(\frac{2\hbar^2 Z^2}{ma_0^2} \left(-\frac{d^2}{dx^2} + \frac{l(l+1)}{x^2} \right) - \frac{2Z^2 e^2}{a_0} \frac{1}{x} \right) u = Eu.\end{aligned}\quad (2.8)$$

Note that

$$\frac{2\hbar^2 Z^2}{ma_0^2} = \frac{2\hbar^2 Z}{ma_0} \frac{me^2}{\hbar^2} = \frac{2Ze^2}{a_0}, \quad (2.9)$$

which reduces our differential equation to

$$\left(-\frac{d^2}{dx^2} + \frac{\ell(\ell+1)}{x^2} - \frac{1}{x} \right) u = \frac{E}{\left(\frac{2Ze^2}{a_0} \right)} u. \quad (2.10)$$

We now define the unit-free parameter κ that encodes the energy:

$$\kappa^2 = -\frac{E}{\left(\frac{2Ze^2}{a_0} \right)} > 0. \quad (2.11)$$

κ is a unit-free version of the bound state energy. The differential equation is then

$$\left(-\frac{d^2}{dx^2} + \frac{\ell(\ell+1)}{x^2} - \frac{1}{x} \right) u = -\kappa^2 u. \quad (2.12)$$

We can further simplify this equation prior to solving it by examining the limiting cases. In the limit $x \rightarrow \infty$, the dominant terms are the second derivative and the term on the right-hand side, giving

$$\frac{d^2 u}{dx^2} = \kappa^2 u \implies u \sim e^{\pm \kappa x}. \quad (2.13)$$

Since κ is unit free, we can make the above exponent equal to a new unit-free coordinate ρ :

$$\boxed{\rho \equiv \kappa x = \frac{2\kappa Z}{a_0} r.} \quad (2.14)$$

This time we get

$$\boxed{\left(-\frac{d^2}{d\rho^2} + \frac{\ell(\ell+1)}{\rho^2} - \frac{1}{\kappa\rho} \right) u = -u.} \quad (2.15)$$

Note that we did not get κ to disappear from the equation. This is good news: the equation should fix the possible values of κ (or possible energies). The equation above is not quite ready for a series solution: we would find a three term recursion relation, which is rather complicated. To make progress we discuss the behavior for small and large ρ .

For $\rho \rightarrow \infty$ we now get $u \sim u^{\pm\rho}$ and of course we hope for $u = e^{-\rho}$ for normalizability. As we discussed before, for $\rho \rightarrow 0$ the radial solution must be of the form $u \sim \rho^{(l+1)}$. This information about the behavior for small and for large ρ suggests a good ansatz for $u(\rho)$

$$\boxed{u(\rho) = \rho^{\ell+1} W(\rho) e^{-\rho}.} \quad (2.16)$$

where $W(\rho)$ is a yet to be determined function that we hope satisfies a simpler differential equation. To derive this differential equation for $W(\rho)$, we plug our ansatz into Eq. (2.15). As a little help on the calculation, we give an intermediate result:

$$-u'' + \frac{\ell(\ell+1)}{\rho^2} u + u = \left(-W'' - \frac{2(\ell+1)}{\rho} W' + \frac{2(\ell+1)}{\rho} W + 2W' \right) \rho^{\ell+1} e^{-\rho}. \quad (2.17)$$

With a little more work we finally get the differential equation for W :

$$\boxed{\rho \frac{d^2W}{d\rho^2} + 2(\ell+1-\rho) \frac{dW}{d\rho} + \left[\frac{1}{\kappa} - 2(\ell+1) \right] W = 0.} \quad (2.18)$$

This looks a bit more complicated than the differential equation we started with but it leads to a very nice one-step recursion relation. As usual we write W as a series expansion

$$W = \sum_{k=0}^{\infty} a_k \rho^k, \quad (2.19)$$

and plugging back into (2.18), we group terms of order ρ^k to derive a recursion relation

$$\begin{aligned} a_{k+1} k(k+1) + 2(\ell+1)(k+1)a_{k+1} - 2ka_k + \left[\frac{1}{\kappa} - 2(\ell+1) \right] a_k &= 0, \\ \rightarrow a_{k+1}(k(k+1) + 2(\ell+1)(k+1)) &= a_k \left(2(\ell+k+1) - \frac{1}{\kappa} \right), \end{aligned} \quad (2.20)$$

which gives

$$\frac{a_{k+1}}{a_k} = \frac{2(k+\ell+1) - \frac{1}{\kappa}}{(k+1)(k+2\ell+2)}. \quad (2.21)$$

Detailed examination shows that for normalizable wave functions, the series must terminate. To see this we examine the large k behavior of the above ratio:

$$\frac{a_{k+1}}{a_k} \simeq \frac{2k}{k^2} = \frac{2}{k}. \quad (2.22)$$

Note that $\frac{2}{k+1} < \frac{2}{k}$; thus if the ratio $\frac{2}{k+1}$ leads to a divergence so will the ratio $\frac{2}{k}$. Taking

$$\frac{a_{k+1}}{a_k} = \frac{2}{k+1} \rightarrow a_{k+1} = \frac{2}{k+1} a_k, \quad (2.23)$$

and this is solved by

$$a_k = \frac{2^k}{k!} a_0. \quad (2.24)$$

Therefore, the sum

$$W = \sum_{k=0}^{\infty} a_k \rho^k \simeq a_0 \sum_{k=0}^{\infty} \frac{2^k \rho^k}{k!} = a_0 e^{2\rho}. \quad (2.25)$$

This is precisely sufficient to make the ansatz in (2.16) un-normalizable.

In order to get a normalizable solution the series for W must terminate. Suppose W is a polynomial of degree N so the coefficients satisfy

$$a_N \neq 0 \quad \text{and} \quad a_{N+1} = 0. \quad (2.26)$$

From Eq.(2.21) this implies

$$\frac{1}{\kappa} = 2(N + \ell + 1). \quad (2.27)$$

Quantization has happened! The energy-encoding parameter κ is now related to integers! Note that ℓ can take values $\ell = 0, 1, 2, \dots$ as it befits it being an angular momentum quantum number. Moreover N can take values $N = 0, 1, 2, \dots$, since a polynomial of degree zero exists, being equal to a constant. Define the **principal** quantum number n as follows:

$$n \equiv N + \ell + 1 = \frac{1}{2\kappa}, \quad \text{with } \ell = 0, 1, 2, \dots, \quad N = 0, 1, 2, \dots, \quad \text{and } n = 1, 2, 3, \dots. \quad (2.28)$$

Importantly, note that for a fixed n we must have

$$0 \leq \ell \leq n - 1, \quad \text{and} \quad 0 \leq N \leq n - 1. \quad (2.29)$$

If n and ℓ are known, N is determined from $N + \ell + 1 = n$. So the independent quantum numbers so far are n and ℓ . Interestingly, the energies depend only on n , since κ depends only on n . Using Eq. (2.11), the energy dependence on principle quantum number is given by

$$E = -\frac{2Z^2 e^2}{a_0} \kappa^2, \quad (2.30)$$

and using $\kappa = \frac{1}{2n}$ we get

$$E = -\frac{Z^2 e^2}{2a_0} \frac{1}{n^2}. \quad (2.31)$$

These are the energy levels of the hydrogen atom! Since at any fixed value of $n > 1$ there are various possible ℓ values, the spectrum is highly degenerate. Even more, each value of ℓ amounts to $2\ell + 1$ states, given the possible values of m . One way to visualize the spectrum is shown in Figure 1. All integer points in the (N, ℓ) positive quadrant represent states. The states with common value of n lie on the dashed lines.

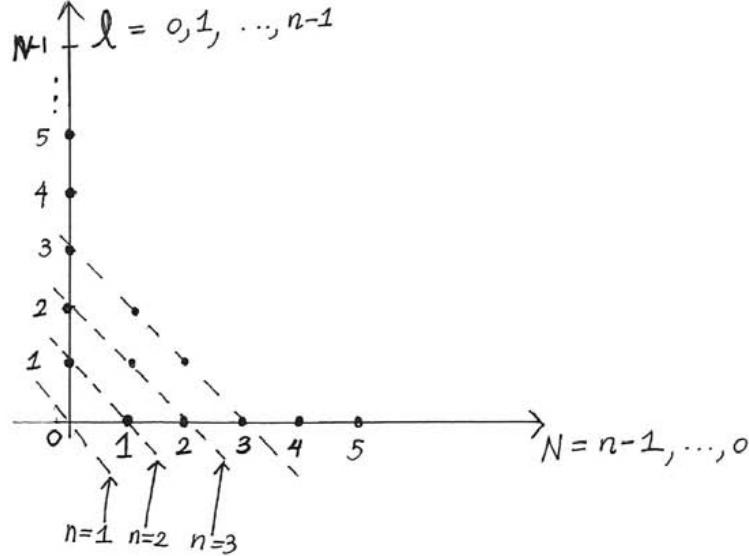


Figure 1: All points with integer $N \geq 0$ and integer $\ell \geq 0$ represent hydrogen atom states. The figure helps us to count the number of possible states for given value of n . Each dot along the diagonal line for a given n represents a possible state.

Figure 1 helps us to count the number of bound states for a given value of n . Recall that for each n , ℓ can take values from $0, \dots, n-1$ and for each value of ℓ , m takes values from $-\ell$ to ℓ . The following table counts the states for the first few values of the principal quantum number n . A given state is specified by its values for (n, ℓ, m) , all of which are referred to as the *quantum numbers* for hydrogen states. Each number has a very important physical meaning: n tells us about the energy eigenvalue, $\hbar^2 \ell(\ell + 1)$ is the eigenvalue of the square of angular momentum and $\hbar m$ is the eigenvalue of the z component of angular momentum.

Number of States

n value	ℓ values	m values	total states
$n=1$	$\ell=0$	$m=0$	1 state
$n=2$	$\ell=0$	$m=0$	1
	$\ell=1$	$m=-1, 0, 1$	+ 3 = 4 states
$n=3$	$\ell=0$	$m=0$	1
	$\ell=1$	$m=-1, 0, 1$	+ 3
	$\ell=2$	$m=-2, \dots, 2$	+ 5 = 9 states

The total number of states for arbitrary principal quantum number n can now be calculated:

$$\# \text{ of states for } n = \sum_{\ell=0}^{n-1} (2\ell + 1) = \frac{2(n-1)n}{2} + n = n^2 - n + n = n^2. \quad (2.32)$$

This is in agreement with the partial results in the table. A more familiar representation of the states of hydrogen is given in Figure 2. The different columns indicate the different values of ℓ . We have also indicated in the figure the values of N , the degree of the polynomial entering the radial solution. Note that for a given ℓ , that is, for a fixed radial equation, the value of N increases as we go up the column. The number N corresponds to the number of nodes in the solution.

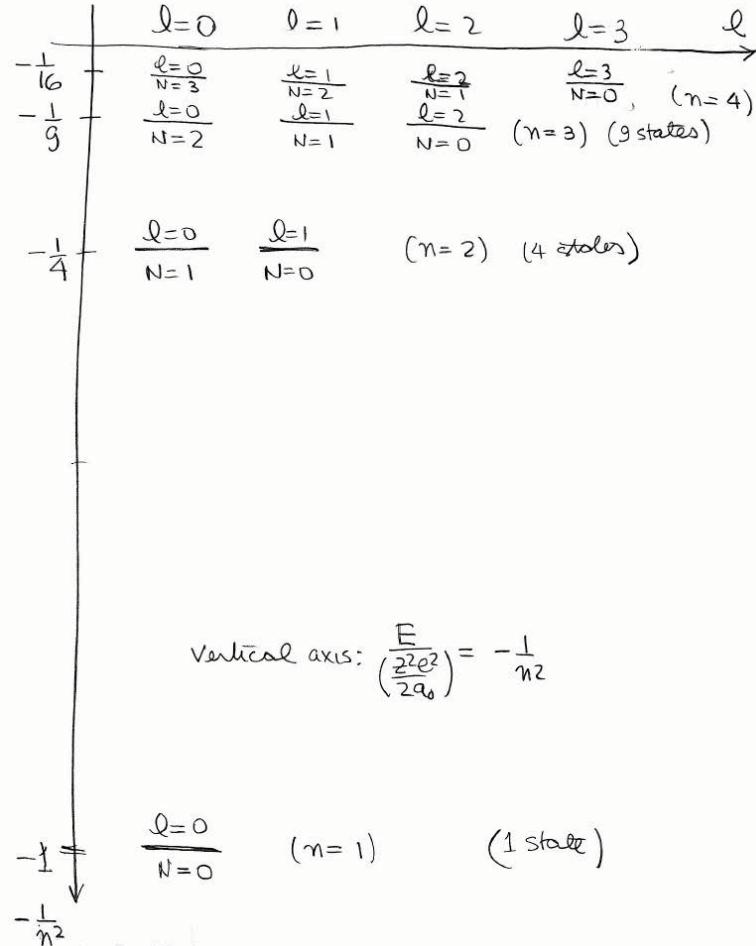


Figure 2: Plot of energy levels $E \sim -1/n^2$ indicating also the angular quantum number ℓ and the degree N of the polynomial. The spectrum is highly degenerate.

Recall that we defined $\rho = \frac{2\kappa Zr}{a_0}$. Together with $\kappa = \frac{1}{2n}$ this gives

$$\boxed{\rho = \frac{Zr}{na_0}.} \quad (2.33)$$

The eigenstates are labelled by quantum numbers (n, ℓ, m) and the wavefunctions are

$$\begin{aligned}\psi_{n\ell m} &= \mathcal{N} \frac{u_{n\ell}(r)}{r} Y_{\ell m}(\theta, \phi) = \mathcal{N} \frac{\rho^{\ell+1}}{\rho} W_{n\ell}(\rho) e^{-\rho} Y_{\ell m}(\theta, \phi) \\ &= \mathcal{N} \rho^\ell \underbrace{W_{n\ell}(\rho)}_{\substack{\text{polynomial of degree} \\ N=n-(\ell+1)}} e^{-\rho} Y_{\ell m}(\theta, \phi),\end{aligned}\tag{2.34}$$

where \mathcal{N} is a normalization constant. Therefore, using the expression for ρ and absorbing constants into \mathcal{N} we have

$$\psi_{n\ell m}(r, \theta, \phi) = \mathcal{N} \left(\frac{r}{a_0} \right)^\ell \left(\begin{array}{l} \text{polynomial in } \frac{r}{a_0} \\ \text{of degree } N=n-(\ell+1) \end{array} \right) e^{-\frac{Zr}{na_0}} Y_{\ell m}(\theta, \phi).\tag{2.35}$$

For the ground state of hydrogen ($Z = 1$), we have $(n, \ell, m) = (1, 0, 0)$. Having zero angular momentum the associated wavefunction has no angular dependence. The normalized wavefunction is

$$\psi_{100}(r, \theta, \phi) = \frac{1}{\sqrt{\pi a_0^3}} e^{-r/a_0}.\tag{2.36}$$

For normalized hydrogen wavefunctions at $n = 2$ and $n = 3$ see

<http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/hydwf.html>

Sarah Geller and Andrew Turner transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare
<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

ANGULAR MOMENTUM

B. Zwiebach
December 16, 2013

Contents

1	Orbital angular momentum and central potentials	1
1.1	Quantum mechanical vector identities	2
1.2	Properties of angular momentum	6
1.3	The central potential Hamiltonian	9
2	Algebraic theory of angular momentum	11
3	Comments on spherical harmonics	18
4	The radial equation	20
5	The free particle and the infinite spherical well	24
5.1	Free particle	24
5.2	The infinite spherical well	25
6	The three-dimensional isotropic oscillator	28
7	Hydrogen atom and Runge-Lenz vector	33

1 Orbital angular momentum and central potentials

Classically the angular momentum vector \vec{L} is defined as the cross-product of the position vector \vec{r} and the momentum vector \vec{p} :

$$\vec{L} = \vec{r} \times \vec{p}. \quad (1.1)$$

In cartesian components, this equation reads

$$\begin{aligned} L_x &= yp_z - zp_y, \\ L_y &= zp_x - xp_z, \\ L_z &= xp_y - yp_x. \end{aligned} \quad (1.2)$$

In quantum mechanics the classical vectors \vec{r} , \vec{p} and \vec{L} become operators. More precisely, they give us triplets of operators:

$$\begin{aligned}\vec{r} &\rightarrow (\hat{x}, \hat{y}, \hat{z}), \\ \vec{p} &\rightarrow (\hat{p}_x, \hat{p}_y, \hat{p}_z), \\ \vec{L} &\rightarrow (\hat{L}_x, \hat{L}_y, \hat{L}_z).\end{aligned}\tag{1.3}$$

When we want more uniform notation, instead of x , y , and z labels we use 1, 2 and 3 labels:

$$\begin{aligned}(\hat{x}_1, \hat{x}_2, \hat{x}_3) &\equiv (\hat{x}, \hat{y}, \hat{z}), \\ (\hat{p}_1, \hat{p}_2, \hat{p}_3) &\equiv (\hat{p}_x, \hat{p}_y, \hat{p}_z), \\ (\hat{L}_1, \hat{L}_2, \hat{L}_3) &\equiv (\hat{L}_x, \hat{L}_y, \hat{L}_z).\end{aligned}\tag{1.4}$$

The basic canonical commutation relations then are easily summarized as

$$[\hat{x}_i, \hat{p}_j] = i\hbar \delta_{ij}, \quad [\hat{x}_i, \hat{x}_j] = 0, \quad [\hat{p}_i, \hat{p}_j] = 0.\tag{1.5}$$

Thus, for example, \hat{x} commutes with \hat{y} , \hat{z} , \hat{p}_y and \hat{p}_z , but fails to commute with \hat{p}_x . In view of (1.2) and (1.3) it is natural to define the angular momentum *operators* by

$$\begin{aligned}\hat{L}_x &\equiv \hat{y}\hat{p}_z - \hat{z}\hat{p}_y, \\ \hat{L}_y &\equiv \hat{z}\hat{p}_x - \hat{x}\hat{p}_z, \\ \hat{L}_z &\equiv \hat{x}\hat{p}_y - \hat{y}\hat{p}_x.\end{aligned}\tag{1.6}$$

Note that these equations are free of ordering ambiguities: each product involves a coordinate and a momentum that commute! In terms of numbered operators

$$\begin{aligned}\hat{L}_1 &\equiv \hat{x}_2\hat{p}_3 - \hat{x}_3\hat{p}_2, \\ \hat{L}_2 &\equiv \hat{x}_3\hat{p}_1 - \hat{x}_1\hat{p}_3, \\ \hat{L}_3 &\equiv \hat{x}_1\hat{p}_2 - \hat{x}_2\hat{p}_1.\end{aligned}\tag{1.7}$$

Note that the angular momentum operators are Hermitian, since \hat{x}_i and \hat{p}_i are and the products can be reordered without cost:

$$\hat{L}_i^\dagger = \hat{L}_i.\tag{1.8}$$

1.1 Quantum mechanical vector identities

We will write triplets of operators as boldfaced vectors, each element of the triplet multiplied by a unit basis vector, just like we do for ordinary vectors. Thus, for example, we have

$$\begin{aligned}\mathbf{r} &\equiv \hat{x}_1\vec{e}_1 + \hat{x}_2\vec{e}_2 + \hat{x}_3\vec{e}_3, \\ \mathbf{p} &\equiv \hat{p}_1\vec{e}_1 + \hat{p}_2\vec{e}_2 + \hat{p}_3\vec{e}_3, \\ \mathbf{L} &\equiv \hat{L}_1\vec{e}_1 + \hat{L}_2\vec{e}_2 + \hat{L}_3\vec{e}_3.\end{aligned}\tag{1.9}$$

These boldface objects are a bit unusual. They are vectors whose components happen to be operators! Moreover, the basis vectors \vec{e}_i must be declared to commute with any of the operators. The boldface objects are useful whenever we want to use the dot products and cross products of three-dimensional space.

Let us, for generality consider vectors \mathbf{a} and \mathbf{b}

$$\begin{aligned}\mathbf{a} &\equiv a_1 \vec{e}_1 + a_2 \vec{e}_2 + a_3 \vec{e}_3, \\ \mathbf{b} &\equiv b_1 \vec{e}_1 + b_2 \vec{e}_2 + b_3 \vec{e}_3,\end{aligned}\tag{1.10}$$

and we will assume that the a_i 's and b_j 's are operators that do not commute. The following are then standard definitions:

$$\begin{aligned}\mathbf{a} \cdot \mathbf{b} &\equiv a_i b_i, \\ (\mathbf{a} \times \mathbf{b})_i &\equiv \epsilon_{ijk} a_j b_k.\end{aligned}\tag{1.11}$$

The order of the operators in the above right-hand sides cannot be changed; it was chosen conveniently, to be the same as the order of the operators on the left-hand sides. We also define,

$$\mathbf{a}^2 \equiv \mathbf{a} \cdot \mathbf{a}.\tag{1.12}$$

Since the operators do not commute, familiar properties of vector analysis do not hold. For example $\mathbf{a} \cdot \mathbf{b}$ is not equal to $\mathbf{b} \cdot \mathbf{a}$. Indeed,

$$\mathbf{a} \cdot \mathbf{b} = a_i b_i = [a_i, b_i] + b_i a_i,\tag{1.13}$$

so that

$$\boxed{\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} + [a_i, b_i].}\tag{1.14}$$

As an application we have

$$\mathbf{r} \cdot \mathbf{p} = \mathbf{p} \cdot \mathbf{r} + [\hat{x}_i, \hat{p}_i],\tag{1.15}$$

The right-most commutator gives $i\hbar \delta_{ii} = 3i\hbar$ so that we have the amusing three-dimensional identity

$$\boxed{\mathbf{r} \cdot \mathbf{p} = \mathbf{p} \cdot \mathbf{r} + 3i\hbar}.\tag{1.16}$$

For cross products we typically have $\mathbf{a} \times \mathbf{b} \neq -\mathbf{b} \times \mathbf{a}$. Indeed,

$$\begin{aligned}(\mathbf{a} \times \mathbf{b})_i &= \epsilon_{ijk} a_j b_k = \epsilon_{ijk} ([a_j, b_k] + b_k a_j) \\ &= -\epsilon_{ikj} b_k a_j + \epsilon_{ijk} [a_j, b_k]\end{aligned}\tag{1.17}$$

where we flipped the k, j indices in one of the epsilon tensors in order to identify a cross product. Indeed, we have now

$$\boxed{(\mathbf{a} \times \mathbf{b})_i = -(\mathbf{b} \times \mathbf{a})_i + \epsilon_{ijk} [a_j, b_k].}\tag{1.18}$$

The simplest example of the use of this identity is one where we use \mathbf{r} and \mathbf{p} . Certainly

$$\mathbf{r} \times \mathbf{r} = 0, \quad \text{and} \quad \mathbf{p} \times \mathbf{p} = 0, \quad (1.19)$$

and more nontrivially,

$$(\mathbf{r} \times \mathbf{p})_i = -(\mathbf{p} \times \mathbf{r})_i + \epsilon_{ijk} [\hat{x}_j, \hat{p}_k]. \quad (1.20)$$

The last term vanishes for it is equal to $i\hbar \epsilon_{ijk} \delta_{jk} = 0$ (the epsilon symbol is antisymmetric in j, k while the delta is symmetric in j, k , resulting in a zero result). We therefore have, quantum mechanically,

$$\boxed{\mathbf{r} \times \mathbf{p} = -\mathbf{p} \times \mathbf{r}.} \quad (1.21)$$

Thus \mathbf{r} and \mathbf{p} can be moved across in the cross product but not in the dot product.

Exercise 1. Prove the following identities for Hermitian conjugation

$$\begin{aligned} (\mathbf{a} \cdot \mathbf{b})^\dagger &= \mathbf{b}^\dagger \cdot \mathbf{a}^\dagger, \\ (\mathbf{a} \times \mathbf{b})^\dagger &= -\mathbf{b}^\dagger \times \mathbf{a}^\dagger. \end{aligned} \quad (1.22)$$

Our definition of the angular momentum operators in (1.7) and the notation developed above imply that we have

$$\boxed{\mathbf{L} = \mathbf{r} \times \mathbf{p} = -\mathbf{p} \times \mathbf{r}.} \quad (1.23)$$

Indeed, given the definition of the product, we have

$$\boxed{\hat{L}_i = \epsilon_{ijk} \hat{x}_j \hat{p}_k.} \quad (1.24)$$

If you write out what this means for $i = 1, 2, 3$ (do it!) you will recover the expressions in (1.7). The angular operator is Hermitian. Indeed, using (1.22) and recalling that \mathbf{r} and \mathbf{p} are Hermitian we have

$$\mathbf{L}^\dagger = (\mathbf{r} \times \mathbf{p})^\dagger = -\mathbf{p}^\dagger \times \mathbf{r}^\dagger = -\mathbf{p} \times \mathbf{r} = \mathbf{L}. \quad (1.25)$$

The use of vector notation implies that, for example,

$$\mathbf{L}^2 = \mathbf{L} \cdot \mathbf{L} = \hat{L}_1 \hat{L}_1 + \hat{L}_2 \hat{L}_2 + \hat{L}_3 \hat{L}_3 = \hat{L}_i \hat{L}_i. \quad (1.26)$$

The classical angular momentum operator is orthogonal to both \vec{r} and \vec{p} as it is built from the cross product of these two vectors. Happily, these properties also hold for the quantum angular momentum. Take for example the dot product of \mathbf{r} with \mathbf{L} to get

$$\mathbf{r} \cdot \mathbf{L} = \hat{x}_i \hat{L}_i = \hat{x}_i \epsilon_{ijk} \hat{x}_j \hat{p}_k = \epsilon_{ijk} \hat{x}_i \hat{x}_j \hat{p}_k = 0. \quad (1.27)$$

The last expression is zero because the \hat{x} 's commute and thus form an object symmetric in i, j , while the epsilon symbol is antisymmetric in i, j . Similarly,

$$\mathbf{p} \cdot \mathbf{L} = \hat{p}_i \hat{L}_i = -\hat{p}_i (\mathbf{p} \times \mathbf{r})_i = -\hat{p}_i \epsilon_{ijk} \hat{p}_j \hat{x}_k = -\epsilon_{ijk} \hat{p}_i \hat{p}_j \hat{x}_k = 0. \quad (1.28)$$

In summary:

$$\boxed{\mathbf{r} \cdot \mathbf{L} = \mathbf{p} \cdot \mathbf{L} = 0.} \quad (1.29)$$

In manipulating vector products the following identity is quite useful

$$\epsilon_{ijk} \epsilon_{ipq} = \delta_{jp} \delta_{kq} - \delta_{jq} \delta_{kp}. \quad (1.30)$$

Its contraction is also needed sometimes:

$$\epsilon_{ijk} \epsilon_{ijq} = 2\delta_{kq}. \quad (1.31)$$

For triple products we find

$$\begin{aligned} [\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_k &= \epsilon_{kji} a_j (\mathbf{b} \times \mathbf{c})_i \\ &= \epsilon_{kji} \epsilon_{ipq} a_j b_p c_q \\ &= -\epsilon_{ijk} \epsilon_{ipq} a_j b_p c_q \\ &= -(\delta_{jp} \delta_{kq} - \delta_{jq} \delta_{kp}) a_j b_p c_q \\ &= a_j b_k c_j - a_j b_j c_k \\ &= [a_j, b_k] c_j + b_k a_j c_j - a_j b_j c_k \\ &= [a_j, b_k] c_j + b_k (\mathbf{a} \cdot \mathbf{c}) - (\mathbf{a} \cdot \mathbf{b}) c_k \end{aligned} \quad (1.32)$$

We can write this as

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b} (\mathbf{a} \cdot \mathbf{c}) - (\mathbf{a} \cdot \mathbf{b}) \mathbf{c} + [a_j, b_k] c_j. \quad (1.33)$$

The first two terms are familiar from vector analysis, the last term is quantum mechanical.

Another familiar relation from vector analysis is the classical

$$(\vec{a} \times \vec{b})^2 \equiv (\vec{a} \times \vec{b}) \cdot (\vec{a} \times \vec{b}) = \vec{a}^2 \vec{b}^2 - (\vec{a} \cdot \vec{b})^2. \quad (1.34)$$

In deriving such an equation you assume that the vector components are commuting numbers, not operators. If we have vector operators additional terms arise.

Exercise 2. Show that

$$\begin{aligned} (\mathbf{a} \times \mathbf{b})^2 &= \mathbf{a}^2 \mathbf{b}^2 - (\mathbf{a} \cdot \mathbf{b})^2 \\ &\quad - a_j [a_j, b_k] b_k + a_j [a_k, b_k] b_j - a_j [a_k, b_j] b_k - a_j a_k [b_k, b_j], \end{aligned} \quad (1.35)$$

and verify that this yields

$$(\mathbf{a} \times \mathbf{b})^2 = \mathbf{a}^2 \mathbf{b}^2 - (\mathbf{a} \cdot \mathbf{b})^2 + \gamma \mathbf{a} \cdot \mathbf{b}, \quad \text{when } [a_i, b_j] = \gamma \delta_{ij}, \quad \gamma \in \mathbb{C}, \quad [b_i, b_j] = 0. \quad (1.36)$$

As an application we calculate \mathbf{L}^2

$$\mathbf{L}^2 = (\mathbf{r} \times \mathbf{p})^2, \quad (1.37)$$

equation (1.36) can be applied with $\mathbf{a} = \mathbf{r}$ and $\mathbf{b} = \mathbf{p}$. Since $[a_i, b_j] = [\hat{x}_i, \hat{p}_j] = i\hbar \delta_{ij}$ we read that $\gamma = i\hbar$, so that

$$\boxed{\mathbf{L}^2 = \mathbf{r}^2 \mathbf{p}^2 - (\mathbf{r} \cdot \mathbf{p})^2 + i\hbar \mathbf{r} \cdot \mathbf{p}.} \quad (1.38)$$

Another useful and simple identity is the following

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}, \quad (1.39)$$

as you should confirm in a one-line computation. In commuting vector analysis this triple product is known to be cyclically symmetric. Note, that in the above no operator has been moved across each other –that's why it holds.

1.2 Properties of angular momentum

A key property of the angular momentum operators is their commutation relations with the \hat{x}_i and \hat{p}_i operators. You should verify that

$$\boxed{\begin{aligned} [\hat{L}_i, \hat{x}_j] &= i\hbar \epsilon_{ijk} \hat{x}_k, \\ [\hat{L}_i, \hat{p}_j] &= i\hbar \epsilon_{ijk} \hat{p}_k. \end{aligned}} \quad (1.40)$$

We say that these equations mean that \mathbf{r} and \mathbf{p} are vectors *under* rotations.

Exercise 3. Use the above relations and (1.18) to show that

$$\mathbf{p} \times \mathbf{L} = -\mathbf{L} \times \mathbf{p} + 2i\hbar \mathbf{p}. \quad (1.41)$$

Hermitization is the process by which we construct a Hermitian operator starting from a non-Hermitian one. Say Ω is not hermitian, its Hermitization Ω_h is defined to be

$$\Omega_h \equiv \frac{1}{2}(\Omega + \Omega^\dagger). \quad (1.42)$$

Exercise 4. Show that the Hermitization of $\mathbf{p} \times \mathbf{L}$ is

$$(\mathbf{p} \times \mathbf{L})_h = \frac{1}{2}(\mathbf{p} \times \mathbf{L} - \mathbf{L} \times \mathbf{p}) = \mathbf{p} \times \mathbf{L} - i\hbar \mathbf{p}. \quad (1.43)$$

Inspired by the behavior of \mathbf{r} and \mathbf{p} under rotations, we declare that an operator \mathbf{u} defined by the triplet $(\hat{u}_1, \hat{u}_2, \hat{u}_3)$ is a *vector under rotations* if

$$[\hat{L}_i, \hat{u}_j] = i\hbar \epsilon_{ijk} \hat{u}_k. \quad (1.44)$$

Exercise 5. Use the above commutator to show that with \mathbf{n} a constant vector,

$$[\mathbf{n} \cdot \mathbf{L}, \mathbf{u}] = -i\hbar \mathbf{n} \times \mathbf{u}. \quad (1.45)$$

Given two operators \mathbf{u} and \mathbf{v} that are vectors under rotations you will show that their dot product is a scalar – it commutes with all \hat{L}_i – and their cross product is a vector:

$$[\hat{L}_i, \mathbf{u} \cdot \mathbf{v}] = 0,$$

$$[\hat{L}_i, (\mathbf{u} \times \mathbf{v})_j] = i\hbar \epsilon_{ijk} (\mathbf{u} \times \mathbf{v})_k.$$

(1.46)

Exercise 6. Prove the above equations.

A number of useful commutator identities follow from (1.46). Most importantly, from the second one, taking $\mathbf{u} = \mathbf{r}$ and $\mathbf{v} = \mathbf{p}$ we get

$$[\hat{L}_i, (\mathbf{r} \times \mathbf{p})_j] = i\hbar \epsilon_{ijk} (\mathbf{r} \times \mathbf{p})_k, \quad (1.47)$$

which gives the celebrated Lie algebra of angular momentum¹

$$[\hat{L}_i, \hat{L}_j] = i\hbar \epsilon_{ijk} \hat{L}_k.$$

(1.48)

More explicitly,

$$\begin{aligned} [\hat{L}_x, \hat{L}_y] &= i\hbar \hat{L}_z, \\ [\hat{L}_y, \hat{L}_z] &= i\hbar \hat{L}_x, \\ [\hat{L}_z, \hat{L}_x] &= i\hbar \hat{L}_y. \end{aligned} \quad (1.49)$$

Note the cyclic nature of these equations: take the first and cycle indices ($x \rightarrow y \rightarrow z \rightarrow x$) to obtain the other two. Another set of examples follows from the first identity in (1.46) using our list of vector operators. For example

$$[\hat{L}_i, \mathbf{r}^2] = [\hat{L}_i, \mathbf{p}^2] = [\hat{L}_i, \mathbf{r} \cdot \mathbf{p}] = 0, \quad (1.50)$$

¹We do not know who celebrated it because we were not invited.

and, very importantly,

$$[\hat{L}_i, \mathbf{L}^2] = 0. \quad (1.51)$$

This equation is the reason the operator \mathbf{L}^2 plays a very important role in the study of central potentials. \mathbf{L}^2 will feature as one of the operators in complete sets of commuting observables. An operator, such as \mathbf{L}^2 , that commutes with all the angular momentum operators is called a “Casimir” of the algebra of angular momentum. Note that the validity of (1.51) just uses the algebra of the \hat{L}_i operators not, for example, how they are built from \mathbf{r} and \mathbf{p} .

Exercise 7. Use (1.18) and the algebra of \hat{L} operators to show that

$$\mathbf{L} \times \mathbf{L} = i\hbar \mathbf{L}. \quad (1.52)$$

This is a very elegant way to express the algebra of angular momentum. In fact, we can show that it is totally equivalent to (1.48). Thus we write

$$\boxed{\mathbf{L} \times \mathbf{L} = i\hbar \mathbf{L} \iff [\hat{L}_i, \hat{L}_j] = i\hbar \epsilon_{ijk} \hat{L}_k.} \quad (1.53)$$

Commutation relations of the form

$$[a_i, b_j] = \epsilon_{ijk} c_k, \quad (1.54)$$

admit a natural rewriting in terms of cross products. From (1.18)

$$(\mathbf{a} \times \mathbf{b})_i + (\mathbf{b} \times \mathbf{a})_i = \epsilon_{ijk} [a_j, b_k] = \epsilon_{ijk} \epsilon_{jkl} c_l = 2 c_i. \quad (1.55)$$

This means that

$$\boxed{[a_i, b_j] = \epsilon_{ijk} c_k \rightarrow \mathbf{a} \times \mathbf{b} + \mathbf{b} \times \mathbf{a} = 2 \mathbf{c}.} \quad (1.56)$$

The arrow does *not* work in the reverse direction. One finds $[a_i, b_j] = \epsilon_{ijk} c_k + s_{ij}$ where $s_{ij} = s_{ji}$ is arbitrary and is *not* determined. If the arrow could be reversed, $\mathbf{a} \times \mathbf{b} + \mathbf{b} \times \mathbf{a} = 0$ would imply that \mathbf{a} and \mathbf{b} commute. We have, however, a familiar example where this does not happen: while $\mathbf{r} \times \mathbf{p} + \mathbf{p} \times \mathbf{r} = 0$ (see (1.23)), the operators \mathbf{r} and \mathbf{p} don't commute.

For a vector \mathbf{u} under rotations, equation (1.56) becomes

$$\mathbf{L} \times \mathbf{u} + \mathbf{u} \times \mathbf{L} = 2i\hbar \mathbf{u}. \quad (1.57)$$

1.3 The central potential Hamiltonian

Angular momentum plays a crucial role in the study of three-dimensional central potential problems. Those are problems where the Hamiltonian describes a particle moving in a potential $V(r)$ that depends just on r , the distance of the particle to the chosen origin. The Hamiltonian takes the form

$$H = \frac{\mathbf{p}^2}{2m} + V(r). \quad (1.58)$$

When writing the Schrödinger equation in position space we identify

$$\mathbf{p} = \frac{\hbar}{i} \nabla, \quad (1.59)$$

and therefore

$$\mathbf{p}^2 = -\hbar^2 \nabla^2, \quad (1.60)$$

where ∇^2 denotes the Laplacian operator –a second order differential operator. In spherical coordinates the Laplacian is well known and gives us

$$\mathbf{p}^2 = -\hbar^2 \left[\frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{r^2} \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right) \right]. \quad (1.61)$$

Our goal is to relate the “angular” part of the above differential operator to angular momentum operators. This will be done by calculating \mathbf{L}^2 and relating it to \mathbf{p}^2 . Since we had from (1.38)

$$\mathbf{L}^2 = \mathbf{r}^2 \mathbf{p}^2 - (\mathbf{r} \cdot \mathbf{p})^2 + i\hbar \mathbf{r} \cdot \mathbf{p}, \quad (1.62)$$

We solve for \mathbf{p}^2 to get

$$\mathbf{p}^2 = \frac{1}{\mathbf{r}^2} \left[(\mathbf{r} \cdot \mathbf{p})^2 - i\hbar \mathbf{r} \cdot \mathbf{p} + \mathbf{L}^2 \right]. \quad (1.63)$$

Let us now consider the above equation in coordinate space, where \mathbf{p} is a gradient. We then have:

$$\mathbf{r} \cdot \mathbf{p} = \frac{\hbar}{i} r \frac{\partial}{\partial r}, \quad (1.64)$$

and thus

$$(\mathbf{r} \cdot \mathbf{p})^2 - i\hbar \mathbf{r} \cdot \mathbf{p} = -\hbar^2 \left(r \frac{\partial}{\partial r} r \frac{\partial}{\partial r} + r \frac{\partial}{\partial r} \right) = -\hbar^2 \left(r^2 \frac{\partial^2}{\partial r^2} + 2r \frac{\partial}{\partial r} \right). \quad (1.65)$$

It then follows that

$$\frac{1}{\mathbf{r}^2} \left[(\mathbf{r} \cdot \mathbf{p})^2 - i\hbar \mathbf{r} \cdot \mathbf{p} \right] = -\hbar^2 \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right) = -\hbar^2 \frac{1}{r} \frac{\partial^2}{\partial r^2} r, \quad (1.66)$$

where the last step is readily checked by explicit expansion. Back in (1.63) we get

$$\mathbf{p}^2 = -\hbar^2 \frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{r^2} \mathbf{L}^2. \quad (1.67)$$

Comparing with (1.61) we identify \mathbf{L}^2 as the operator

$$\boxed{\mathbf{L}^2 = -\hbar^2 \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right).} \quad (1.68)$$

Note that the units are fully carried by the \hbar^2 in front and that the differential operator is completely angular: it has no radial dependence. Given our expression (1.67) for \mathbf{p}^2 we can now rewrite the three-dimensional Hamiltonian as

$$\boxed{H = \frac{\mathbf{p}^2}{2m} + V(r) = -\frac{\hbar^2}{2m} \frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{2mr^2} \mathbf{L}^2 + V(r).} \quad (1.69)$$

A key property of central potential problems is that the angular momentum operators commute with the Hamiltonian

$$\text{Central potential Hamiltonians: } [\hat{L}_i, H] = 0. \quad (1.70)$$

We have seen that \hat{L}_i commutes with \mathbf{p}^2 so it is only needed to show that the \hat{L}_i commute with $V(r)$. This is eminently reasonable, for \hat{L}_i commutes with $\mathbf{r}^2 = r^2$, so one would expect it to commute with any function of $\sqrt{\mathbf{r}^2} = r$. In the problem set you will consider this question and develop a formal argument that confirms the expectation.

The above commutator implies that the \hat{L}_i operators are conserved in central potentials. Indeed

$$i\hbar \frac{d}{dt} \langle \hat{L}_i \rangle = \langle [\hat{L}_i, H] \rangle = 0. \quad (1.71)$$

We can now consider the issue of complete sets of commuting observables. The list of operators that we have is

$$H, \hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{L}_1, \hat{L}_2, \hat{L}_3, \mathbf{r}^2, \mathbf{p}^2, \mathbf{r} \cdot \mathbf{p}, \mathbf{L}^2, \dots \quad (1.72)$$

where, for the time being, we included all operators up to squares of coordinates, momenta, and angular momenta. Since we want to understand the spectrum of the Hamiltonian, one of the labels of states will be the energy and thus H must be in the list of commuting observables. Because of the potential $V(r)$ none of the \hat{p}_i operators commute the Hamiltonian. Because of the \mathbf{p}^2 term in the Hamiltonian none of the \hat{x}_i commute with the Hamiltonian. Nor will $\mathbf{r}^2, \mathbf{p}^2$ and $\mathbf{r} \cdot \mathbf{p}$. The list is thus reduced to

$$H, \hat{L}_1, \hat{L}_2, \hat{L}_3, \mathbf{L}^2. \quad (1.73)$$

where there are no dots anymore, since without \hat{x}_i or \hat{p}_i there are no other operators to build (recall also that $\mathbf{L} \times \mathbf{L} = i\hbar\mathbf{L}$ and thus it is not new). All the operators in the list commute

with H : the \hat{L}_i as discussed in (1.70), and \mathbf{L}^2 because, after all, it is built from \hat{L}_i . But all the operators do not commute with each other. From the \hat{L}_i we can only pick at most one, for then the other two necessarily do not commute with the chosen one. Happily we can also keep \mathbf{L}^2 because of its Casimir property (1.51). Conventionally, everybody chooses $\hat{L}_3 = \hat{L}_z$ as one element of the set of commuting observables. Thus we have

$$\boxed{\text{Commuting observables: } H, \hat{L}_z, \mathbf{L}^2.} \quad (1.74)$$

We can wonder if this set is complete in the sense that all energy eigenstates are uniquely labelled by the eigenvalues of the above operators. The answer is yes, for the bound state spectrum of a particle that has no other degrees of freedom (say, no spin).

2 Algebraic theory of angular momentum

Hermitian operators $\hat{J}_x, \hat{J}_y, \hat{J}_z$ are said to satisfy the algebra of **angular momentum** if the following commutation relations:

$$[\hat{J}_i, \hat{J}_j] = i\hbar \epsilon_{ijk} \hat{J}_k. \quad (2.1)$$

More explicitly, in components

$$\begin{aligned} [\hat{J}_x, \hat{J}_y] &= i\hbar \hat{J}_z \\ [\hat{J}_y, \hat{J}_z] &= i\hbar \hat{J}_x \\ [\hat{J}_z, \hat{J}_x] &= i\hbar \hat{J}_y. \end{aligned} \quad (2.2)$$

The \hat{J}_i operators could be \hat{L}_i , or \hat{S}_i or something else! Will only use this algebra and the Hermiticity of the operators. From this algebra it also follows that

$$[\hat{J}_i, \mathbf{J}^2] = 0. \quad (2.3)$$

This can be checked explicitly, but our proof of the analogous result (1.51): $[\hat{L}_i, \mathbf{L}^2] = 0$ only used the algebra of the operators \hat{L}_i , so this also holds for the \hat{J}_i , which satisfy the same algebra. It is not convenient to define

$$\boxed{\begin{aligned} \hat{J}_+ &\equiv \hat{J}_x + i\hat{J}_y, \\ \hat{J}_- &\equiv \hat{J}_x - i\hat{J}_y, \end{aligned}} \quad (2.4)$$

such that the two operators are Hermitian conjugates of each other:

$$(\hat{J}_+)^{\dagger} = \hat{J}_-. \quad (2.5)$$

Note that both \hat{J}_x and \hat{J}_y can be solved for in terms of \hat{J}_+ and \hat{J}_- . It is useful to compute the algebra of the operators \hat{J}_+ , \hat{J}_- , and \hat{J}_z . We begin by computing the product $\hat{J}_+ \hat{J}_-$:

$$\hat{J}_+ \hat{J}_- = \hat{J}_x^2 + \hat{J}_y^2 - i[\hat{J}_x, \hat{J}_y] = \hat{J}_x^2 + \hat{J}_y^2 + \hbar \hat{J}_z. \quad (2.6)$$

Together with the product in the opposite order we have

$$\begin{aligned} \hat{J}_+ \hat{J}_- &= \hat{J}_x^2 + \hat{J}_y^2 + \hbar \hat{J}_z, \\ \hat{J}_- \hat{J}_+ &= \hat{J}_x^2 + \hat{J}_y^2 - \hbar \hat{J}_z. \end{aligned} \quad (2.7)$$

From these two we can quickly get the commutator:

$$[\hat{J}_+, \hat{J}_-] = 2\hbar \hat{J}_z. \quad (2.8)$$

Moreover, we obtain two expressions for $\hat{J}_x^2 + \hat{J}_y^2$

$$\hat{J}_x^2 + \hat{J}_y^2 = \hat{J}_+ \hat{J}_- - \hbar \hat{J}_z = \hat{J}_- \hat{J}_+ + \hbar \hat{J}_z. \quad (2.9)$$

Adding \hat{J}_z^2 to both sides of the equation we find

$$\boxed{\mathbf{J}^2 = \hat{J}_+ \hat{J}_- + \hat{J}_z^2 - \hbar \hat{J}_z = \hat{J}_- \hat{J}_+ + \hat{J}_z^2 + \hbar \hat{J}_z.} \quad (2.10)$$

Of course, since \hat{J}_i and \mathbf{J}^2 commute, we also have

$$[\hat{J}_\pm, \mathbf{J}^2] = 0. \quad (2.11)$$

We finally have to compute the commutator of \hat{J}_\pm with \hat{J}_z . This is quickly done:

$$[\hat{J}_z, \hat{J}_+] = [\hat{J}_z, \hat{J}_x] + i[\hat{J}_z, \hat{J}_y] = i\hbar \hat{J}_y + i(-i\hbar \hat{J}_x) = \hbar(\hat{J}_x + i\hat{J}_y) = \hbar \hat{J}_+. \quad (2.12)$$

Similarly, $[\hat{J}_z, \hat{J}_-] = -\hbar \hat{J}_-$ and therefore, all in all

$$\boxed{[\hat{J}_z, \hat{J}_\pm] = \pm \hbar \hat{J}_\pm.} \quad (2.13)$$

This is similar to our harmonic oscillator commutators $[\hat{N}, \hat{a}^\dagger] = \hat{a}^\dagger$ and $[\hat{N}, \hat{a}] = -\hat{a}$, if we identify \hat{N} with \hat{J}_z , \hat{a}^\dagger with \hat{J}_+ and \hat{a} with \hat{J}_- . In the oscillator case we learned from these that, acting on states, \hat{a}^\dagger raises the \hat{N} eigenvalue by one unit while \hat{a} decreases it by one unit. As we will see, \hat{J}_+ adds \hbar to the \hat{J}_z eigenvalue and \hat{J}_- subtracts \hbar to the \hat{J}_z eigenvalue.

Since \mathbf{J}^2 and \hat{J}_z are hermitian and commute, they can be simultaneously diagonalized. In fact, there are no more operators in the angular momentum algebra can be added to this list of simultaneously diagonalizable operators. The common eigenstates form an orthonormal basis

for the relevant vector space. We thus introduce eigenstates $|j, m\rangle$, with $j, m \in \mathbb{R}$, where the first label relates to the \mathbf{J}^2 eigenvalue and the second label to the \hat{J}_z eigenvalue:

$$\begin{aligned}\mathbf{J}^2|j, m\rangle &= \hbar^2 j(j+1) |j, m\rangle, \\ \hat{J}_z |j, m\rangle &= \hbar m |j, m\rangle.\end{aligned}\tag{2.14}$$

The orthonormality of states implies that

$$\langle j', m' | j, m \rangle = \delta_{j', j} \delta_{m', m},\tag{2.15}$$

where we assumed that we will not have to deal with continuous values of j, m that would require delta function normalization. This will be confirmed below. Since j and m are real, the eigenvalues of the hermitian operators are real, as they have to be. The first line shows that the eigenvalue of \mathbf{J}^2 is defined to be $\hbar^2 j(j+1)$. This can seem curious: why not $\hbar^2 j^2$? The answer is convenience, as we will see below. Alternatively, if we know $\hbar^2 j(j+1)$, how do we get j ? For this first note that $\hbar^2 j(j+1)$ must be non-negative:

$$\hbar^2 j(j+1) = \langle j, m | \mathbf{J}^2 | j, m \rangle = \sum_{i=1}^3 \langle j, m | \hat{J}_i \hat{J}_i | j, m \rangle = \sum_{i=1}^3 \| \hat{J}_i | j, m \rangle \|^2 \geq 0,\tag{2.16}$$

where in the first step we used the eigenvalue definition and orthonormality. Therefore the condition

$$j(j+1) \geq 0\tag{2.17}$$

is the only a priori condition on the values of j . Since what matters is the eigenvalue of \mathbf{J}^2 we can use any of the two j 's that give a particular value of $j(j+1)$. As shown in the figure below, the positivity of $j(j+1)$ requires $j \geq 0$ or $j \leq -1$. We can simply use $j \geq 0$.

$$\text{States are labeled as } |j, m\rangle \text{ with } j \geq 0.\tag{2.18}$$

You should not think that there are two different states, with two different j 's associated with the eigenvalue $\hbar j(j+1)$. It is just one state, that we are labeling in an unusual way. Of course, a theory may end up having more than one state with the same \mathbf{J}^2 eigenvalue. In that case we will have more than one state with the same $j > 0$.

Let us now investigate what the operators \hat{J}_{\pm} do when acting on the above eigenstates. Since they commute with \mathbf{J}^2 , the operators J_+ or J_- do not change the j value of a state:

$$\mathbf{J}^2(\hat{J}_{\pm}|j, m\rangle) = \hat{J}_{\pm} \mathbf{J}^2 |j, m\rangle = \hbar j(j+1)(\hat{J}_{\pm}|j, m\rangle),\tag{2.19}$$

so that we must have

$$J_{\pm}|j, m\rangle \propto |j, m'\rangle, \text{ for some } m'.\tag{2.20}$$

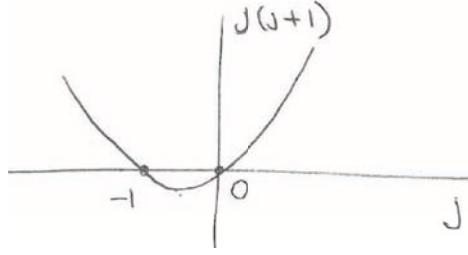


Figure 1: Since $j(j + 1) \geq 0$ for consistency, we can label the states $|j, m\rangle$ using $j \geq 0$.

On the other hand, as anticipated above, the \hat{J}_\pm operators change the value of m :

$$\begin{aligned}\hat{J}_z(\hat{J}_\pm|j, m\rangle) &= ([\hat{J}_z, J_\pm] + J_\pm \hat{J}_z)|j, m\rangle \\ &= (\pm \hbar J_\pm + \hbar m J_\pm)|j, m\rangle \\ &= \hbar(m \pm 1)(J_\pm|j, m\rangle),\end{aligned}\tag{2.21}$$

from which we learn that

$$\hat{J}_\pm|j, m\rangle = C_\pm(j, m)|j, m \pm 1\rangle,\tag{2.22}$$

where $C_\pm(j, m)$ is a constant to be determined. Indeed, we see that \hat{J}_+ raised the m eigenvalue by one unit while \hat{J}_- decreases the m eigenvalue by one unit. To determine $C_\pm(j, m)$ we first take the adjoint of the above equation

$$\langle j, m|\hat{J}_\mp = \langle j, m \pm 1|C_\pm(j, m)^*,\tag{2.23}$$

and then form the overlap

$$\langle j, m|J_\mp J_\pm|j, m\rangle = |C_\pm(j, m)|^2.\tag{2.24}$$

To evaluate the left-hand side use (2.10) in the form

$$\hat{J}_\mp \hat{J}_\pm = \mathbf{J}^2 - \hat{J}_z^2 \mp \hbar \hat{J}_z\tag{2.25}$$

as well as $\langle j, m|j, m\rangle = 1$:

$$|C_\pm(j, m)|^2 = \langle j, m|(\mathbf{J}^2 - \hat{J}_z^2 \mp \hbar \hat{J}_z)|j, m\rangle = \hbar^2 j(j + 1) - \hbar^2 m^2 \mp \hbar^2 m.\tag{2.26}$$

We have thus found that

$$|C_\pm(j, m)|^2 = \hbar^2 (j(j + 1) - m(m \pm 1)) = \|\hat{J}_\pm|j, m\rangle\|^2.\tag{2.27}$$

Here we learn a few things. If we start with a consistent state $|j, m\rangle$ of norm one (as assumed above), the states $\hat{J}_\pm|j, m\rangle \sim |j, m \pm 1\rangle$ created by the action of \hat{J}_\pm on $|j, m\rangle$ are inconsistent if

the middle expression in the above relation is negative. This is because that middle expression is in fact the norm-squared of $\hat{J}_\pm|j, m\rangle$. Assuming that middle expression is positive (or zero) we can take $C_\pm(j, m)$ real and equal to the positive square root

$$C_\pm(j, m) = \hbar\sqrt{j(j+1) - m(m \pm 1)}. \quad (2.28)$$

We have thus obtained

$$J_\pm|j, m\rangle = \hbar\sqrt{j(j+1) - m(m \pm 1)} |j, m \pm 1\rangle.$$

(2.29)

Given a consistent state $|j, m\rangle$, how far can we raise or lower the value of m ? Our classical intuition is that $|\hat{J}_z| \leq |\mathbf{J}|$. So we should get something like $|m| \lesssim \sqrt{j(j+1)}$.

Consider this in two steps:

1. For the raised state to be consistent we must have $\|J_+|j, m\rangle\|^2 \geq 0$ and therefore

$$j(j+1) - m(m+1) \geq 0 \rightarrow m(m+1) \leq j(j+1) \quad (2.30)$$

The solution to the inequality is given in figure 2:

$$-j-1 \leq m \leq j. \quad (2.31)$$

If we had chosen $\hbar^2 j(j+1)$ to be the eigenvalue of \mathbf{J}^2 this inequality would not have had a simple solution.

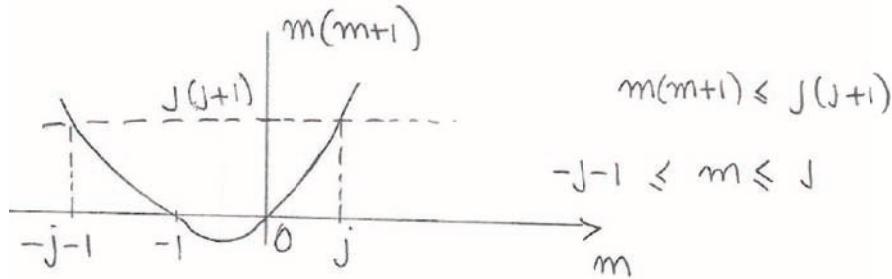


Figure 2: Solving the inequality $m(m+1) \leq j(j+1)$.

Since we are raising m we can focus on the troubles that raising can give given that $m \leq j$. Assume $m = j - \beta$ with $0 < \beta < 1$ so that the inequality (2.31) is satisfied and m is less than one unit below j . Then the raising once gives us a state with $m' = m + 1 > j$ and since the inequality is now violated raising one more time would then give an inconsistent

state. To prevent such inconsistency the process of raising must terminate: there must be a state that raising gives no state (the zero state). That indeed happens only if $m = j$ since then $C_+(j, j) = 0$

$$\hat{J}_+|j, j\rangle = 0. \quad (2.32)$$

2. For the lowered state to be consistent we must have $|\hat{J}_-|j, m\rangle|^2 \geq 0$ and therefore

$$j(j+1) - m(m-1) \geq 0 \rightarrow m(m-1) \leq j(j+1) \quad (2.33)$$

The solution to this inequality is obtained using figure 3 and gives

$$-j \leq m \leq j+1. \quad (2.34)$$

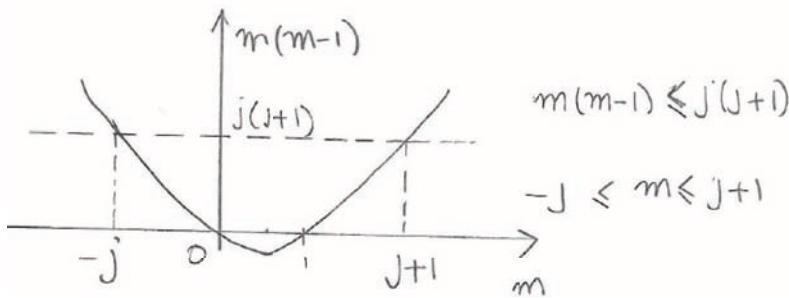


Figure 3: Solving the inequality $m(m-1) \leq j(j+1)$.

This time we can focus here on $m \geq -j$ and the complications due to lowering. Assume $m = -j + \beta$ with $0 < \beta < 1$ so that the constraint (2.34) is satisfied and m is less than one unit above $-j$. Then lowering once gives us a state with $m' = m - 1 < -j$ and since the inequality is now violated lowering one more time would then give an inconsistent state. To prevent such inconsistency we need that lowering terminates on some state for which lowering gives no state (the zero state). That indeed happens only if $m = -j$ since then $C_-(j, -j) = 0$

$$\hat{J}_-|j, -j\rangle = 0. \quad (2.35)$$

The above analysis shows that for consistency a multiplet of states with some given fixed j must be such that the m values must include $-j$ and $+j$. Since m is increased or decreased by integer steps via the \hat{J}_\pm operators, the distance $2j$ between j and $-j$ must be an integer:

$$2j \in \mathbb{Z} \rightarrow j \in \mathbb{Z}/2, \rightarrow j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots \quad (2.36)$$

This is the fundamental quantization of angular momentum. Angular momentum can be integral or half-integral. For any allowed value of j the m values will be $j, j-1, \dots, -j$. Thus the multiplet with angular momentum j has the following $2j+1$ states

$$\begin{aligned} & |j, j\rangle, \\ & |j, j-1\rangle, \\ & \vdots \\ & |j, -j\rangle. \end{aligned} \tag{2.37}$$

For $j = 0$ there is just one state, the **singlet** with $m = 0$: $|0, 0\rangle$.

For $j = \frac{1}{2}$ we have two states:

$$\begin{aligned} & |\frac{1}{2}, \frac{1}{2}\rangle, \\ & |\frac{1}{2}, -\frac{1}{2}\rangle. \end{aligned} \tag{2.38}$$

These are the states of a spin-1/2 particle, when we identify the angular momentum \mathbf{J} with the spin angular momentum operator \mathbf{S} . The top state has $\hat{S}_z = \hbar/2$ and the lower state has $\hat{S}_z = -\hbar/2$. These are our conventional $|+\rangle$ and $|-\rangle$ states, respectively.

For $j = 1$ we have three states :

$$\begin{aligned} & |1, 1\rangle, \\ & |1, 0\rangle, \\ & |1, -1\rangle. \end{aligned} \tag{2.39}$$

For $j = 3/2$ we have four states:

$$\begin{aligned} & |\frac{3}{2}, \frac{3}{2}\rangle \\ & |\frac{3}{2}, \frac{1}{2}\rangle \\ & |\frac{3}{2}, -\frac{1}{2}\rangle \\ & |\frac{3}{2}, -\frac{3}{2}\rangle \end{aligned} \tag{2.40}$$

One last one! For $j = 2$ we have five states:

$$\begin{aligned} & |2, 2\rangle, \\ & |2, 1\rangle, \\ & |2, 0\rangle, \\ & |2, -1\rangle, \\ & |2, -2\rangle. \end{aligned} \tag{2.41}$$

On any state of a multiplet with angular momentum j the eigenvalue J^2 of \mathbf{J}^2 is

$$J^2 = \hbar^2 j(j+1) \rightarrow \frac{1}{\hbar} J = \sqrt{j(j+1)} \tag{2.42}$$

In the limit as j is large

$$\frac{1}{\hbar} J = j \sqrt{1 + \frac{1}{j}} \simeq j + \frac{1}{2} + \mathcal{O}(1/j). \quad (2.43)$$

So for large j the angular momentum is roughly $J \simeq j$.

3 Comments on spherical harmonics

We have constructed the \mathbf{L}^2 operator as a differential operator in position space, with coordinates r, θ, ϕ . The operator happens to depend only on θ and ϕ and takes the form (1.68)

$$\mathcal{L}^2 = -\hbar^2 \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right). \quad (3.1)$$

where we denoted it with a calligraphic symbol to make it clear we are talking about a differential operator. We also have, with the same notation

$$\hat{\mathcal{L}}_z = \frac{\hbar}{i} \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right). \quad (3.2)$$

A short calculation, passing to spherical coordinates (do it!) shows that

$$\hat{\mathcal{L}}_z = \frac{\hbar}{i} \frac{\partial}{\partial \phi}. \quad (3.3)$$

Finally, a longer calculation shows that

$$\mathcal{L}_\pm = \hbar e^{\pm i\phi} \left(i \cot \theta \frac{\partial}{\partial \phi} \pm \frac{\partial}{\partial \theta} \right). \quad (3.4)$$

Recall now how things work for coordinate representations. For a single coordinate x we had that the operator \hat{p} can be taken out of the matrix element as the differential operator \mathfrak{p} :

$$\langle x | \hat{p} | \psi \rangle = \mathfrak{p} \langle x | \psi \rangle, \text{ where } \mathfrak{p} = \frac{\hbar}{i} \frac{\partial}{\partial x} \quad (3.5)$$

We will let $\langle \theta \phi |$ denote position states on the unit sphere and the spherical harmonic $Y_{\ell m}$ will be viewed as the wavefunction for the state $|\ell m\rangle$ so that

$$Y_{\ell m}(\theta, \phi) \equiv \langle \theta \phi | \ell, m \rangle. \quad (3.6)$$

Consider now (2.14) in the form

$$\begin{aligned} \mathbf{L}^2 |\ell, m\rangle &= \hbar^2 \ell(\ell+1) |\ell, m\rangle, \\ \hat{\mathcal{L}}_z |\ell, m\rangle &= \hbar m |\ell, m\rangle. \end{aligned} \quad (3.7)$$

Letting the bra $\langle \theta\phi|$ act on them we have

$$\begin{aligned}\langle \theta\phi | \mathbf{L}^2 | \ell, m \rangle &= \hbar^2 \ell(\ell+1) \langle \theta\phi | \ell, m \rangle, \\ \langle \theta\phi | \hat{L}_z | \ell, m \rangle &= \hbar m \langle \theta\phi | \ell, m \rangle.\end{aligned}\quad (3.8)$$

Using the analog of (3.5) for our operators we have

$$\begin{aligned}\mathcal{L}^2 \langle \theta\phi | \ell, m \rangle &= \hbar^2 \ell(\ell+1) \langle \theta\phi | \ell, m \rangle, \\ \hat{\mathcal{L}}_z \langle \theta\phi | \ell, m \rangle &= \hbar m \langle \theta\phi | \ell, m \rangle.\end{aligned}\quad (3.9)$$

These are equivalent to

$$\begin{aligned}\mathcal{L}^2 Y_{\ell m}(\theta, \phi) &= \hbar^2 \ell(\ell+1) Y_{\ell m}(\theta, \phi), \\ \hat{\mathcal{L}}_z Y_{\ell m}(\theta, \phi) &= \hbar m Y_{\ell m}(\theta, \phi).\end{aligned}\quad (3.10)$$

where \mathcal{L}^2 and $\hat{\mathcal{L}}_z$ are the coordinate representation operators for \mathbf{L}^2 and \hat{L}_z respectively.

On the unit sphere the measure of integration is $\sin \theta d\theta d\phi$ so we postulate that the completeness relation for the $|\theta\phi\rangle$ position states reads

$$\int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\phi |\theta\phi\rangle \langle \theta\phi| = \mathbf{1}. \quad (3.11)$$

The integral will be written more briefly as

$$\int d\Omega |\theta\phi\rangle \langle \theta\phi| = \mathbf{1} \quad (3.12)$$

where

$$\int d\Omega = \int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\phi = - \int_1^{-1} d(\cos \theta) \int_0^{2\pi} d\phi = \int_{-1}^1 d(\cos \theta) \int_0^{2\pi} d\phi. \quad (3.13)$$

Our orthogonality relation

$$\langle \ell', m' | \ell, m \rangle = \delta_{\ell', l} \delta_{m', m}, \quad (3.14)$$

gives, by including a complete set of position states

$$\int d\Omega \langle \ell', m' | \theta\phi \rangle \langle \theta\phi | \ell, m \rangle = \delta_{\ell', l} \delta_{m', m}. \quad (3.15)$$

This gives the familiar orthogonality property of the spherical harmonics:

$$\int d\Omega Y_{\ell' m'}^*(\theta, \phi) Y_{\ell m}(\theta, \phi) = \delta_{\ell', l} \delta_{m', m}. \quad (3.16)$$

Note that the equation

$$\hat{\mathcal{L}}_z Y_{\ell m} = \hbar m Y_{\ell m} \quad (3.17)$$

together with (3.3) implies that

$$Y_{\ell m}(\theta, \phi) = P_{\ell m}(\theta) e^{im\phi}. \quad (3.18)$$

The ϕ dependence of the spherical harmonics is very simple indeed!

One can show that for spherical harmonics, which are related to **orbital** angular momentum, one can only have integer ℓ . While j can be half-integral, any attempt to define spherical harmonics for half-integral ℓ fails. You will indeed show in the homework that this is necessarily the case.

4 The radial equation

Recall that from (1.69) we have

$$H = \frac{\mathbf{p}^2}{2m} + V(r) = -\frac{\hbar^2}{2m} \frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{2mr^2} \mathcal{L}^2 + V(r) \quad (4.1)$$

where we used the differential operator realization \mathcal{L}^2 of the operator \mathbf{L}^2 . The Schrödinger equation will be solved using the following ansatz for energy eigenstates

$$\Psi_{E\ell m}(\mathbf{x}) = f_{E\ell m}(r) Y_{\ell m}(\theta, \phi) \quad (4.2)$$

We have the product of a radial function $f_{E\ell m}(r)$ times an angular function $Y_{\ell m}$ which is an eigenstate of \mathbf{L}^2 and of \hat{L}_z :

$$\mathcal{L}^2 Y_{\ell m} = \hbar^2 \ell(\ell+1) Y_{\ell m}, \quad \hat{L}_z Y_{\ell m} = \hbar m Y_{\ell m} \quad (4.3)$$

Plugging this into the Schrödinger equation $H\Psi = E\Psi$, the $Y_{\ell m}$ dependence can be cancelled out and we get

$$-\frac{\hbar^2}{2m} \frac{1}{r} \frac{d^2}{dr^2} (rf_{E\ell m}) + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} f_{E\ell m} + V(r)f_{E\ell m} = Ef_{E\ell m} \quad (4.4)$$

We note that this equation does not depend on the quantum number m (do not confuse this with the mass m !) Therefore the label m is not needed in the radial function and we let $f_{E\ell m} \rightarrow f_{E\ell}$ so that we have

$$\Psi_{E\ell m}(\mathbf{x}) = f_{E\ell}(r) Y_{\ell m}(\theta, \phi) \quad (4.5)$$

and the differential equation, multiplying by r becomes

$$-\frac{\hbar^2}{2m} \frac{d^2}{dr^2} (rf_{E\ell}) + \left(V(r) + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} \right) (rf_{E\ell}) = E(rf_{E\ell}) \quad (4.6)$$

This suggests writing introducing a modified radial function $u_{E\ell}(r)$ by

$$f_{E\ell}(r) = \frac{u_{E\ell}(r)}{r} \quad (4.7)$$

so that we have

$$\boxed{\Psi_{E\ell m}(\mathbf{x}) = \frac{u_{E\ell}(r)}{r} Y_{\ell m}(\theta, \phi),} \quad (4.8)$$

with radial equation

$$-\frac{\hbar^2}{2m} \frac{d^2 u_{E\ell}}{dr^2} + V_{\text{eff}}(r) u_{E\ell} = E u_{E\ell}, \quad (4.9)$$

where the effective potential V_{eff} constructed by adding to the potential $V(r)$ the centrifugal barrier term proportional to \mathbf{L}^2 :

$$V_{\text{eff}}(r) \equiv V(r) + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} \quad (4.10)$$

This is like a one-dimensional Schrödinger equation in the variable r , but as opposed to our usual problems with $x \in (-\infty, \infty)$, the radius $r \in [0, \infty]$ and we will need some special care for $r = 0$.

The normalization of our wavefunctions proceeds as follows

$$\int d^3\mathbf{x} |\Psi_{E\ell m}(\mathbf{x})|^2 = 1 \quad (4.11)$$

This gives

$$\int r^2 dr d\Omega \frac{|u_{E\ell}(r)|^2}{r^2} Y_{\ell m}^*(\Omega) Y_{\ell m}(\Omega) = 1 \quad (4.12)$$

the angular integral gives one and we get

$$\int_0^\infty dr |u_{E\ell}(r)|^2 = 1 \quad (4.13)$$

a rather natural result for the function $u_{E\ell}$ that plays the role of radial wavefunction.

Behavior of solutions as $r \rightarrow 0$. We claim that

$$\lim_{r \rightarrow 0} u_{E\ell}(r) = 0. \quad (4.14)$$

This requirement does not arise from normalization: as you can see in (4.13) a finite $u_{E\ell}$ at $r = 0$ would cause no trouble. Imagine taking a solution u_{E0} with $\ell = 0$ that approaches a constant as $r \rightarrow 0$:

$$\lim_{r \rightarrow 0} u_{E0}(r) = c \neq 0. \quad (4.15)$$

The full solution Ψ near the origin would then take the form

$$\Psi(\mathbf{x}) \simeq \frac{c}{r} Y_{00} = \frac{c'}{r}, \quad (4.16)$$

since Y_{00} is simply a constant. The problem with this wavefunction is that it simply *does not solve* the Schrödinger equation! You may remember from electromagnetism that the Laplacian of the $1/r$ function has a delta function at the origin, so that as a result

$$\nabla^2 \Psi(\mathbf{x}) = -4\pi c' \delta(\mathbf{x}). \quad (4.17)$$

Since the Laplacian is part of the Hamiltonian, this delta function must be cancelled by some other contribution, but there is none, since the potential $V(r)$ does not have delta functions².

We can learn more about the behavior of the radial solution under the reasonable assumption that the *centrifugal barrier dominates the potential as $r \rightarrow 0$* . In this case the most singular terms of the radial differential equation must cancel each other out, leaving less singular terms that we can ignore in this leading term calculation. So we set:

$$-\frac{\hbar^2}{2m} \frac{d^2 u_{E\ell}}{dr^2} + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} u_{E\ell} = 0, \quad \text{as } r \rightarrow 0. \quad (4.18)$$

or equivalently

$$\frac{d^2 u_{E\ell}}{dr^2} = \frac{\ell(\ell+1)}{r^2} u_{E\ell}, \quad (4.19)$$

The solutions of this can be taken to be $u_{E\ell} = r^s$ with s a constant to be determined. We then find

$$s(s-1) = \ell(\ell+1) \rightarrow s = \ell+1, s = -\ell \quad (4.20)$$

thus leading to two possible behaviors near $r = 0$:

$$u_{E\ell} \sim r^{\ell+1}, \quad u_{E\ell} \sim \frac{1}{r^\ell}. \quad (4.21)$$

For $\ell = 0$ the second behavior was shown to be inconsistent with the Schrödinger equation at $r = 0$ (because of a delta function). For $\ell > 0$ the second behavior is not consistent with normalization. Therefore we have established that

$$u_{E\ell} \sim c r^{\ell+1}, \quad \text{as } r \rightarrow 0.$$

(4.22)

Note that the full radial dependence is obtained by dividing by r , so that

$$f_{E\ell} \sim c r^\ell, \quad (4.23)$$

This allows for a constant non-zero wavefunction at the origin only for $\ell = 0$. Only for $\ell = 0$ a particle can be at the origin. For $\ell \neq 0$ the angular momentum “barrier” prevents the particle from reaching the origin.

²Delta function potentials in more than one dimension are very singular and require regulation.

Behavior of solutions as $r \rightarrow \infty$. Again, we can make some definite statements once we assume some properties of the potential. Let us consider the case when the potential $V(r)$ vanishes beyond some radius or at least decays fast enough as the radius grows without bound

$$V(r) = 0, \text{ for } r > r_0, \text{ or } \lim_{r \rightarrow \infty} rV(r) = 0. \quad (4.24)$$

Curiously, the above assumptions are violated for the $1/r$ potential of the Hydrogen atom (an extended discussion of related issues can be found in Shankar around page 344). Under these assumptions we ignore the effective potential completely (including the centrifugal barrier) and the equation becomes

$$-\frac{\hbar^2}{2m} \frac{d^2 u_{E\ell}}{dr^2} = Eu_{E\ell}(r). \quad (4.25)$$

The equation is the familiar

$$\frac{d^2 u_{E\ell}}{dr^2} = -\frac{2mE}{\hbar^2} u_{E\ell}. \quad (4.26)$$

The resulting $r \rightarrow \infty$ behavior follows immediately

$$\begin{aligned} E < 0, \quad u_{E\ell} &\sim \exp\left(-\sqrt{\frac{2m|E|}{\hbar^2}} r\right), \\ E > 0, \quad u_{E\ell} &\sim \exp(\pm ikr), \quad k = \sqrt{\frac{2mE}{\hbar^2}}. \end{aligned} \quad (4.27)$$

The first behavior, for $E < 0$ is typical of bound states. For $E > 0$ we have a continuous spectrum with degenerate solutions (hence the \pm). Having understood the behavior of solutions near $r = 0$ and for $r \rightarrow \infty$ this allows for qualitative plots of radial solutions.

The discrete spectrum is organized as follows. We have energy eigenstates for all values of ℓ . In fact for each value of ℓ the potential V_{eff} in the radial equation is different. So this equation must be solved for $\ell = 0, 1, \dots$. For each fixed ℓ we have a one-dimensional problem, so we have no degeneracies in the bound state spectrum. We have a set of allowed values of energies that depend on ℓ and are numbered using an integer $n = 1, 2, \dots$. For each allowed energy $E_{n\ell}$ we have a single radial solution $u_{n\ell}$.

$$\text{Fixed } \ell, \text{ Energies: } E_{n\ell}, \quad \text{Radial function: } u_{n\ell}, \quad n = 1, 2, \dots \quad (4.28)$$

Of course each solution $u_{n\ell}$ for the radial equation represents $2\ell + 1$ degenerate solutions to the Schrödinger equation corresponding to the possible values of \hat{L}_z in the range $(-\ell\hbar, \ell\hbar)$. Note that n has replaced the label E in the radial solution, and the energies have now been labeled. This is illustrated in the diagram of Figure 4, where each solution of the radial equation is shown as a short line atop an ℓ label on the horizontal axis. This is the spectral diagram for the central-potential Hamiltonian. Each line of a given ℓ represents the $(2\ell + 1)$ degenerate states obtained with $m = -\ell, \dots, \ell$. Because the bound state spectrum of a one-dimensional potential

is non-degenerate, our radial equation can't have any degeneracies for any fixed ℓ . Thus the lines on the diagram are single lines! Of course, other types of degeneracies of the spectrum can exist: some states having different values of ℓ may have the same energy. In other words, the states may match across columns on the figure. Finally, note that since the potential becomes more positive as ℓ is increased, the lowest energy state occurs for $\ell = 0$ and the energy $E_{1,\ell}$ of the lowest state for a given ℓ increases as we increase ℓ .

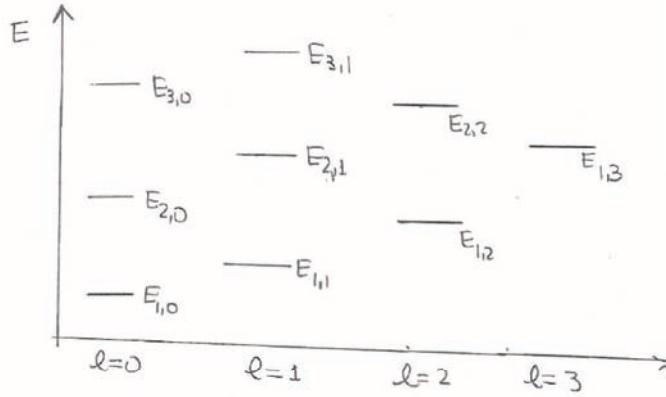


Figure 4: The generic discrete spectrum of a central-potential Hamiltonian, showing the angular momentum ℓ multiplets and their energies.

5 The free particle and the infinite spherical well

5.1 Free particle

It may sound funny at first, but it is interesting to find the radial solutions that correspond to a free particle! A particle that moves in $V(r) = 0$. This amounts to a very different description of the energy eigenstates. In cartesian coordinates we would write solutions as momentum eigenstates, for all values of the momentum. To label such solutions we could use three labels: the components of the momentum. Alternatively, we can use the energy and the direction defined by the momentum, which uses two labels. Here the solutions will be labeled by the energy and (ℓ, m) , the usual two integers that describe the angular dependence (of course, ℓ affects the radial dependence too). The radial equation is

$$-\frac{\hbar^2}{2m} \frac{d^2 u_{E\ell}}{dr^2} + \frac{\hbar^2 \ell(\ell+1)}{2m r^2} u_{E\ell} = Eu_{E\ell} \quad (5.29)$$

which is, equivalently

$$-\frac{d^2 u_{E\ell}}{dr^2} + \frac{\ell(\ell+1)}{r^2} u_{E\ell} = k^2 u_{E\ell}, \quad k \equiv \sqrt{\frac{2mE}{\hbar^2}} \quad (5.30)$$

In this equation there is no quantization of the energy. Indeed we can redefine the radial coordinate in a way that the energy does not appear, namely, k does not appear. Letting $\rho = kr$ the equation becomes

$$-\frac{d^2 u_{E\ell}}{d\rho^2} + \frac{\ell(\ell+1)}{\rho^2} u_{E\ell} = u_{E\ell}, \quad (5.31)$$

The solution of this differential equation with regular behavior at the origin is $u_{E\ell} = c\rho j_\ell(\rho)$, where c is an arbitrary constant. This means that we can take

$$u_{E\ell} = r j_\ell(kr). \quad (5.32)$$

Here the $j_\ell(x)$ are the spherical Bessel functions. All in all we have

$$\text{Free particle: } \Psi_{E\ell m}(\mathbf{x}) = j_\ell(kr) Y_{lm}(\theta, \phi). \quad (5.33)$$

The spherical Bessel functions have the following behavior

$$x j_\ell(x) \sim \frac{x^{\ell+1}}{(2\ell+1)!!}, \text{ as } x \rightarrow 0, \quad \text{and} \quad x j_\ell(x) \sim \sin\left(x - \frac{\ell\pi}{2}\right) \text{ as } x \rightarrow \infty. \quad (5.34)$$

which implies the correct behavior for $u_{E\ell}$ as $r \rightarrow 0$ and $r \rightarrow \infty$. Indeed, for $r \rightarrow \infty$ we have

$$\text{Free particle : } u_{E\ell} \sim \sin\left(kr - \frac{\ell\pi}{2}\right), \text{ as } r \rightarrow \infty. \quad (5.35)$$

Whenever the potential is not zero, but vanishes beyond some radius, the solutions, for $r \rightarrow \infty$ take the form

$$u_{E\ell} \sim \sin\left(kr - \frac{\ell\pi}{2} + \delta_\ell(E)\right), \text{ as } r \rightarrow \infty. \quad (5.36)$$

Here $\delta_\ell(E)$ is called the **phase shift** and by definition vanishes if there is no potential. The form of the solution above is consistent with our general behavior, as this is a superposition of the two solutions available in (4.27) for $E > 0$. The phase shift contains all the information about a potential $V(r)$ available to someone probing the potential from far away by scattering particles off of it.

5.2 The infinite spherical well

An infinite spherical well of radius a is a potential that forces the particle to be within the sphere $r = a$. The potential is zero for $r \leq a$ and it is infinite for $r > a$.

$$V(r) = \begin{cases} 0, & \text{if } r \leq a \\ \infty, & \text{if } r > a \end{cases} \quad (5.37)$$

The Schrödinger radial equation is the same as the one for the free particle

$$-\frac{d^2 u_{E\ell}}{d\rho^2} + \frac{\ell(\ell+1)}{\rho^2} u_{E\ell} = u_{E\ell}, \quad \rho = kr \quad (5.38)$$

where k again encodes the energy E , which is greater than zero. It follows that the solutions are the ones we had before, with spherical Bessel functions, but this time quantization of the energy arises because the wavefunctions must vanish for $r = a$.

Let us do the case $\ell = 0$ without resorting to the Bessel functions. The above equation becomes

$$-\frac{d^2 u_{E,0}}{d\rho^2} = u_{E,0} \quad \rightarrow \quad u_{E,0} = A \sin \rho + B \cos \rho. \quad (5.39)$$

Since the solution must vanish at $r = 0$ we must choose the sin function:

$$u_{E,0}(r) = \sin kr. \quad (5.40)$$

Since this must vanish for $r = a$ we have that k must take values k_n with

$$k_n a = n\pi, \quad \text{for } n = 1, 2, \dots \infty. \quad (5.41)$$

Those values of k_n correspond to energies $E_{n,0}$ where the n indexes the solutions and the 0 represents $\ell = 0$:

$$E_{n,0} = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2}{2ma^2} (k_n a)^2 = \frac{\hbar^2}{2ma^2} n^2 \pi^2. \quad (5.42)$$

Note that $\frac{\hbar^2}{2ma^2}$ is the natural energy scale for this problem and therefore it is convenient to define the unit-free scaled energies $\mathcal{E}_{n,\ell}$ by dividing $E_{n,\ell}$ by the natural energy by

$$\mathcal{E}_{n,\ell} \equiv \frac{2ma^2}{\hbar^2} E_{n,\ell}. \quad (5.43)$$

It follows that the ‘energies’ for $\ell = 0$ are

$$\mathcal{E}_{n,0} = n^2 \pi^2, \quad u_{n,0} = \sin\left(\frac{n\pi r}{a}\right). \quad (5.44)$$

We have

$$\mathcal{E}_{1,0} \simeq 9.8696, \quad \mathcal{E}_{2,0} \simeq 39.478, \quad \mathcal{E}_{3,0} \simeq 88.826, \quad (5.45)$$

Let us now do $\ell = 1$. Here the solutions are $u_{E,1} = r j_1(kr)$. This Bessel function is

$$j_1(\rho) = \frac{\sin \rho}{\rho^2} - \frac{\cos \rho}{\rho} \quad (5.46)$$

The zeroes of $j_1(\rho)$ occur for $\tan \rho = \rho$. Of course, we are not interested in the zero at $\rho = 0$. You can check that the first three zeroes occur for 4.4934, 7.7252, 10.904. For higher values of ℓ it becomes a bit more complicated but there are tables of zeroes on the web.

There is notation in which the nontrivial zeroes are denoted by $z_{n,\ell}$ where

$$z_{n,\ell} \text{ is the } n\text{-th zero of } j_\ell : j_\ell(z_{n,\ell}) = 0. \quad (5.47)$$

The vanishing condition at $r = a$ quantizes k so that

$$k_{n,\ell} a = z_{n,\ell} \quad (5.48)$$

and the energies

$$E_{n,\ell} = \frac{\hbar^2}{2ma^2} (k_{n,\ell}a)^2 \rightarrow \mathcal{E}_{n,\ell} = z_{n,\ell}^2 \quad (5.49)$$

We have

$$\begin{aligned} z_{1,1} &= 4.4934, & z_{2,1} &= 7.7252, & z_{3,1} &= 10.904 \\ z_{1,2} &= 5.7634, & z_{2,2} &= 9.095, \\ z_{1,3} &= 6.9879, & z_{2,3} &= 10.417. \end{aligned} \quad (5.50)$$

which give us

$$\begin{aligned} \mathcal{E}_{1,1} &= 20.191, & \mathcal{E}_{2,1} &= 59.679, & \mathcal{E}_{3,1} &= 118.89 \\ \mathcal{E}_{1,2} &= 33.217, & \mathcal{E}_{2,2} &= 82.719, \\ \mathcal{E}_{1,3} &= 48.83, & \mathcal{E}_{2,3} &= 108.51. \end{aligned} \quad (5.51)$$

The main point to be made is that there are no accidental degeneracies: the energies for different values of ℓ never coincide. More explicitly, with $\ell \neq \ell'$ we have that $\mathcal{E}_{n,\ell} \neq \mathcal{E}_{n',\ell'}$ for any choices of n and n' . This is illustrated in figure 5.

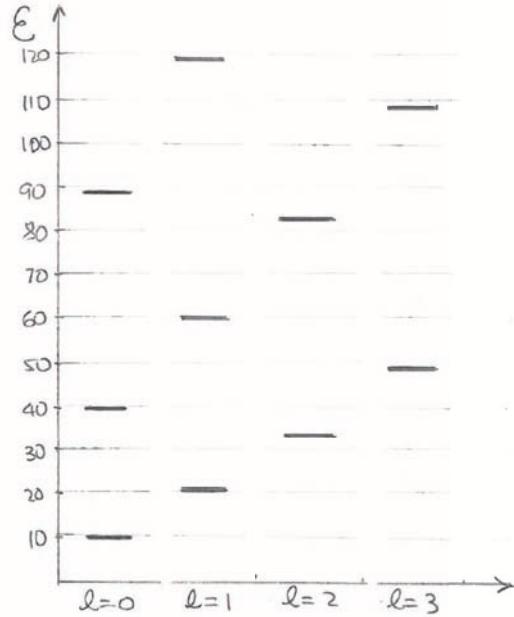


Figure 5: The spectrum of the infinite spherical square well. There are no accidental degeneracies.

6 The three-dimensional isotropic oscillator

The potential of the 3D isotropic harmonic oscillator is as follows:

$$V = \frac{1}{2}m\omega^2(x^2 + y^2 + z^2) = \frac{1}{2}m\omega^2r^2. \quad (6.52)$$

As we will see, the spectrum for this quantum mechanical system has degeneracies, that are explained by the existence of some **hidden symmetry**, a symmetry that is not obvious from the start. Thus in some ways this quantum 3D oscillator is a lot more symmetric than the infinite spherical well.

As you know, for the 3D oscillator we can use creation and annihilation operators $\hat{a}_x^\dagger, \hat{a}_y^\dagger, \hat{a}_z^\dagger$ and $\hat{a}_x, \hat{a}_y, \hat{a}_z$ associated with 1D oscillators in the x, y , and z directions. The Hamiltonian then takes the form:

$$H = \hbar\omega(\hat{N}_1 + \hat{N}_2 + \hat{N}_3 + \frac{3}{2}) = \hbar\omega(\hat{N} + \frac{3}{2}). \quad (6.53)$$

where we defined $\hat{N} \equiv \hat{N}_1 + \hat{N}_2 + \hat{N}_3$.

We now want to explain how tensor products are relevant to the 3D oscillator. We have discussed tensor products before to describe two particles, each associated with a vector space and the combined system associated with the tensor product of vector spaces. But tensor products are also relevant to single particles, if they have degrees of freedom that live in different spaces, or more than one set of attributes, each of which described by states in some vector space. For example, if a spin 1/2 particle can move, the relevant states live in the tensor product of momentum space and the 2-dimensional complex vector space of spin. States are obtained by superposition of basic states of the form $|p\rangle \otimes (\alpha|+\rangle + \beta|-\rangle)$

For the 3D oscillator, the Hamiltonian is the sum of commuting Hamiltonians of 1D oscillators for the x, y , and z directions. Thus the general states are obtained by tensoring the state spaces $\mathcal{H}_x, \mathcal{H}_y$, and \mathcal{H}_z of the three independent oscillators. It is a single particle oscillating, but the description of what it is doing entails saying what is doing in each of the independent directions. Thus we write

$$\mathcal{H}_{3D} = \mathcal{H}_x \otimes \mathcal{H}_y \otimes \mathcal{H}_z. \quad (6.54)$$

Instead of this tensor product reflecting the behavior of three different particles, this tensor product allows us to describe the behavior of one particle in three different directions. The vacuum state $|0\rangle$ of the 3D oscillator can be viewed as

$$|0\rangle \equiv |0\rangle_x \otimes |0\rangle_y \otimes |0\rangle_z. \quad (6.55)$$

The associated wavefunction is

$$\Psi(x, y, z) = \langle x| \otimes \langle y| \otimes \langle z| |0\rangle = \langle x|0\rangle_x \langle y|0\rangle_y \langle z|0\rangle_z = \psi_0(x)\psi_0(y)\psi_0(z). \quad (6.56)$$

where ψ_0 is the ground state wavefunction of the 1D oscillator. This is the expected answer. Recalling the form of (non-normalized) basis states for \mathcal{H}_x , \mathcal{H}_y , and \mathcal{H}_z :

$$\begin{aligned} \text{basis states for } \mathcal{H}_x : & (\hat{a}_x^\dagger)^{n_x} |0\rangle_x, n_x = 0, 1, \dots \\ \text{basis states for } \mathcal{H}_y : & (\hat{a}_y^\dagger)^{n_y} |0\rangle_y, n_y = 0, 1, \dots \\ \text{basis states for } \mathcal{H}_z : & (\hat{a}_z^\dagger)^{n_z} |0\rangle_z, n_z = 0, 1, \dots \end{aligned} \quad (6.57)$$

We then have that the basis states for the 3D state space are

$$\text{basis states of } \mathcal{H}_{3D} : (\hat{a}_x^\dagger)^{n_x} |0\rangle_x \otimes (\hat{a}_y^\dagger)^{n_y} |0\rangle_y \otimes (\hat{a}_z^\dagger)^{n_z} |0\rangle_z, n_x, n_y, n_z \in \{0, 1, \dots\} \quad (6.58)$$

This is what we would expect intuitively, we simply pile arbitrary numbers of \hat{a}_x^\dagger , \hat{a}_y^\dagger , and \hat{a}_z^\dagger on the vacuum. It is this multiplicative structure that is the signature of tensor products. Having understood the above, for brevity we write such basis states simply as

$$(\hat{a}_x^\dagger)^{n_x} (\hat{a}_y^\dagger)^{n_y} (\hat{a}_z^\dagger)^{n_z} |0\rangle. \quad (6.59)$$

Each of the states in (6.58) has a wavefunction that is the product of x , y , and z -dependent wavefunctions. Once we form superpositions of such states, the total wavefunction cannot any longer be factorized into x , y , and z -dependent wavefunctions. The x , y , and z -dependences become ‘entangled’. These are precisely the analogs of entangled states of three particles.

We are ready to begin constructing the individual states of the 3D isotropic harmonic oscillator system. The key property is that the states must organize themselves into representations of angular momentum. Since angular momentum commutes with the Hamiltonian, angular momentum multiplets represent degenerate states.

We already built the ground state, which is a single state with \hat{N} eigenvalue $N = 0$. All other states have higher energies, so this state must be, by itself a representation of angular momentum. It can only be the singlet $\ell = 0$. Thus we have

$$N = 0, E = \frac{3}{2}\hbar\omega, |0\rangle \leftrightarrow \ell = 0. \quad (6.60)$$

The states with $N = 1$ have $E = (5/2)\hbar\omega$ and are

$$\hat{a}_x^\dagger |0\rangle, \hat{a}_y^\dagger |0\rangle, \hat{a}_z^\dagger |0\rangle \quad (6.61)$$

These three states fit precisely into an $\ell = 1$ multiplet (a triplet). There is no other possibility, in fact: any higher ℓ multiplet has too many states and we only have 3 degenerate ones. Moreover, we cannot have three singlets, this is a degeneracy inconsistent with the lack of degeneracy for 1D bound states (as discussed earlier). The $\ell = 0$ ground state and the $\ell = 1$ triplet at the first excited level are indicated in Figure 7.

Let us proceed now with the states at $N = 2$ or $E = (7/2)\hbar\omega$. These are, the following six states:

$$(\hat{a}_x^\dagger)^2|0\rangle, (\hat{a}_y^\dagger)^2|0\rangle, (\hat{a}_z^\dagger)^2|0\rangle, \hat{a}_x^\dagger\hat{a}_y^\dagger|0\rangle, \hat{a}_x^\dagger\hat{a}_z^\dagger|0\rangle, \hat{a}_y^\dagger\hat{a}_z^\dagger|0\rangle. \quad (6.62)$$

To help ourselves in trying to find the angular momentum multiplets recall that that the number of states $\#$ for a given ℓ are

ℓ	$\#$
0	1
1	3
2	5
3	7
4	9
5	11
6	13
7	15

Since we cannot use the triplet twice, the only way to get six states is having five from $\ell = 2$ and one from $\ell = 0$. Thus

$$\text{Six } N = 2 \text{ states : } (\ell = 2) \oplus (\ell = 0). \quad (6.63)$$

Note that here we use the direct sum (not the tensor product!) the six states define a six dimensional vector space spanned by five vectors in $\ell = 0$ and one vector in $\ell = 0$. Had we used a tensor product we would just have 5 vectors.

Let us continue to figure out the pattern. At $N = 3$ with $E = (9/2)\hbar\omega$ we actually have 10 states (count them!) It would seem now that there are two possibilities for multiplets

$$(\ell = 3) \oplus (\ell = 1) \text{ or } (\ell = 4) \oplus (\ell = 0) \quad (6.64)$$

We can argue that the second possibility cannot be. The problem with it is that the $\ell = 3$ multiplet, which has not appeared yet, would not arise at this level. If it would arise later, it would do so at a higher energy, and we would have the lowest $\ell = 3$ multiplet above the lowest $\ell = 4$ multiplet, which is not possible. You may think that perhaps $\ell = 3$ multiplets never appear and the inconsistency is avoided, but this is not true. At any rate we will give below a more rigorous argument. The conclusion, however is that

$$\text{Ten } N = 3 \text{ states : } (\ell = 3) \oplus (\ell = 1). \quad (6.65)$$

Let us do the next level! At $N = 4$ we find 15 states. Instead of writing them out let us count them without listing them. In fact, we can easily do the general case of arbitrary integer $N \geq 1$. The states we are looking for are of the form

$$(\hat{a}_x^\dagger)^{n_x}(\hat{a}_y^\dagger)^{n_y}(\hat{a}_z^\dagger)^{n_z}|0\rangle, \text{ with } n_x + n_y + n_z = N. \quad (6.66)$$

We need to count how many different solutions there are to $n_x + n_y + n_z = N$, with $n_x, n_y, n_z \geq 0$. This is the number of states $\#(N)$ at level N . To visualize this think of $n_x + n_y + n_z = N$ as the equation for a plane in three-dimensional space with axes n_x, n_y, n_z . Since no integer can be negative, we are looking for points with integer coordinates in the region of the plane that lies on the positive octant, as shown in Figure 6. Starting at one of the three corners, say $(n_x, n_y, n_z) = (N, 0, 0)$ we have one point, then moving towards the origin we encounter two points, then three, and so on until we find $N+1$ points on the (n_y, n_z) plane. Thus, the number of states $\#(N)$ for number N is

$$\#(N) = 1 + 2 + \dots + (N+1) = \frac{(N+1)(N+2)}{2} \quad (6.67)$$

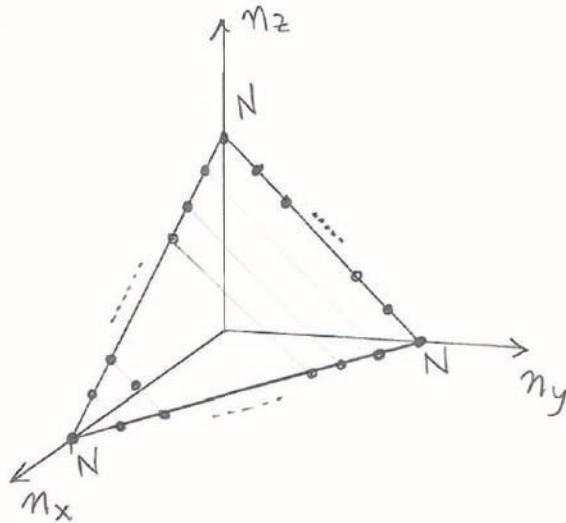


Figure 6: Counting the number of degenerate states with number N in the 3D simple harmonic oscillator.

Back to the $N = 4$ level, $\#(4)=15$. We rule out a single $\ell = 7$ multiplet since states with $\ell = 4, 5, 6$ have not appeared yet. By this logic the highest ℓ multiplet for $N = 4$ must be the lowest that has not appeared yet, thus $\ell = 4$, with 9 states. The remaining six must appear as $\ell = 2$ plus $\ell = 0$. Thus, we have

$$15 \text{ } N = 4 \text{ states : } (\ell = 4) \oplus (\ell = 2) \oplus (\ell = 0) . \quad (6.68)$$

Thus we see that ℓ jumps by steps of two, starting from the maximal ℓ . This is in fact the rule. It is quickly confirmed for the $\#(5)=21$ states with $N = 5$ would arise from $(\ell = 5) \oplus (\ell = 3) \oplus (\ell = 1)$. All this is shown in Figure 7.

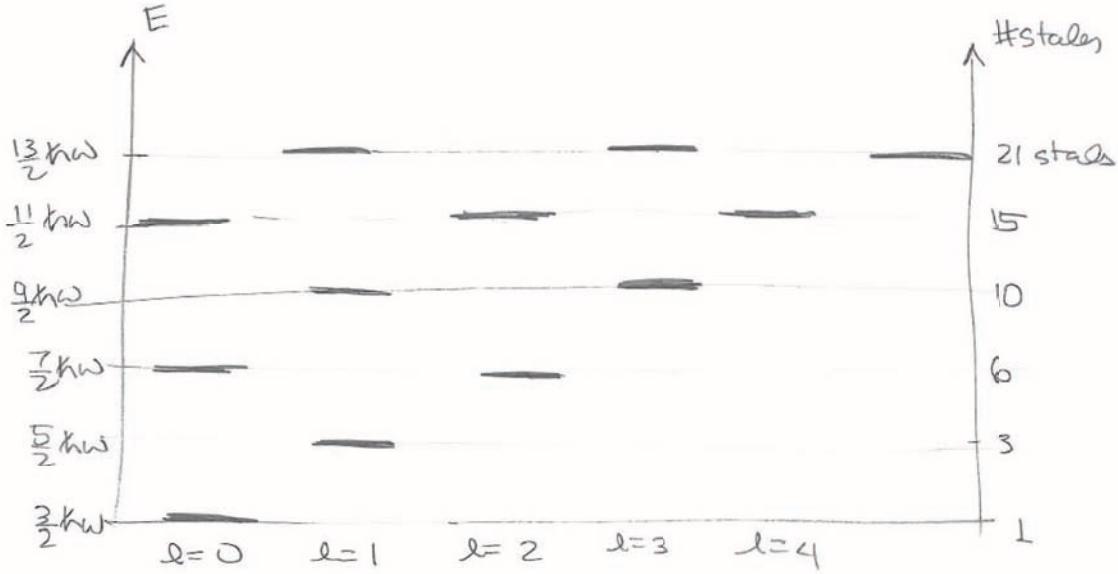


Figure 7: Spectral diagram for angular momentum multiplets in the 3D isotropic harmonic oscillator.

Some of the structure of angular momentum multiplets can be seen more explicitly using the \hat{a}_L and \hat{a}_R operators introduced for the 2D harmonic oscillator:

$$\hat{a}_L = \frac{1}{\sqrt{2}}(\hat{a}_x + i\hat{a}_y), \quad \hat{a}_R = \frac{1}{\sqrt{2}}(\hat{a}_x - i\hat{a}_y). \quad (6.69)$$

L and R objects commute with each other and we have $[\hat{a}_L, \hat{a}_L^\dagger] = [\hat{a}_R, \hat{a}_R^\dagger] = 1$. With number operators $\hat{N}_R = \hat{a}_R^\dagger \hat{a}_R$ and $\hat{N}_L = \hat{a}_L^\dagger \hat{a}_L$ we then have $H = \hbar\omega(\hat{N}_R + \hat{N}_L + \hat{N}_z + \frac{3}{2})$ and, more importantly, the z component \hat{L}_z of angular momentum takes the simple form

$$\hat{L}_z = \hbar(\hat{N}_R - \hat{N}_L). \quad (6.70)$$

Note that \hat{a}_z carries no z -component of angular momentum. States are now built acting with arbitrary numbers of \hat{a}_L^\dagger , \hat{a}_R^\dagger and \hat{a}_z^\dagger operators on the vacuum. The $N = 1$ states are then presented as

$$\hat{a}_R^\dagger |0\rangle, \quad \hat{a}_z^\dagger |0\rangle, \quad \hat{a}_L^\dagger |0\rangle. \quad (6.71)$$

We see that the first state has $L_z = \hbar$, the second $L_z = 0$ and the third $L_z = -\hbar$, exactly the three expected values of the $\ell = 1$ multiplet identified before. For number $N = 2$ the state with highest L_z is $(\hat{a}_R^\dagger)^2 |0\rangle$ and it has $L_z = 2\hbar$. This shows that the highest ℓ multiplet is $\ell = 2$. For arbitrary positive integer number N , the state with highest L_z is $(\hat{a}_R^\dagger)^N |0\rangle$ and it has $L_z = \hbar N$. This shows we must have an $\ell = N$ multiplet. This is in fact what we got before! We can also understand the reason for the jump of two units from the top state of the multiplet. Consider

the above state with maximal \hat{L}_z/\hbar equal to N and then the states with one and two units less of \hat{L}_z/\hbar :

$$\begin{aligned}\hat{L}_z/\hbar = N &: (\hat{a}_R^\dagger)^N |0\rangle \\ \hat{L}_z/\hbar = N - 1 &: (\hat{a}_R^\dagger)^{N-1} \hat{a}_z^\dagger |0\rangle \\ \hat{L}_z/\hbar = N - 2 &: (\hat{a}_R^\dagger)^{N-2} (\hat{a}_z^\dagger)^2 |0\rangle, \quad (\hat{a}_R^\dagger)^{N-1} \hat{a}_L^\dagger |0\rangle\end{aligned}\tag{6.72}$$

While there is only one state with one unit less of \hat{L}_z/\hbar there are two states with two units less. One linear combination of these two states must belong to the $\ell = N$ multiplet, but the other linear combination must be the top state of an $\ell = N - 2$ multiplet! This is the reason for the jump of two units.

For arbitrary N we can see why $\#(N)$ can be reproduced by ℓ multiplets skipping by two

$$\begin{aligned}N \text{ odd : } \#(N) &= \underbrace{1 + 2}_{\ell=1} + \underbrace{3 + 4}_{\ell=3} + \underbrace{5 + 6}_{\ell=5} + \underbrace{7 + 8}_{\ell=7} + \dots + \underbrace{N + (N+1)}_{\ell=N} \\ N \text{ even : } \#(N) &= \underbrace{1}_{\ell=0} + \underbrace{2 + 3}_{\ell=2} + \underbrace{4 + 5}_{\ell=4} + \underbrace{6 + 7}_{\ell=6} + \dots + \underbrace{N + (N+1)}_{\ell=N}\end{aligned}\tag{6.73}$$

The accidental degeneracy is “explained” if we identify an operator that commutes with the Hamiltonian (a symmetry) and connects the various ℓ multiplets that appear for a fixed number N . One such operator is

$$K \equiv \hat{a}_R^\dagger \hat{a}_L .\tag{6.74}$$

You can check it commutes with the Hamiltonian, and with a bit more work, that acting on the top state of the $\ell = N - 2$ multiplet it gives the top state of the $\ell = N$ multiplet.

7 Hydrogen atom and Runge-Lenz vector

The hydrogen atom Hamiltonian is

$$H = \frac{\mathbf{p}^2}{2m} - \frac{e^2}{r} .\tag{7.75}$$

The natural length scale here is the Bohr radius a_0 , which is the unique length that can be built using the constants in this Hamiltonian: \hbar , m , and e^2 . We determine a_0 by setting $p \sim \hbar/a_0$ and equating magnitudes of kinetic and potential terms, ignoring numerical factors:

$$\frac{\hbar^2}{ma_0^2} = \frac{e^2}{a_0} \rightarrow a_0 = \frac{\hbar^2}{me^2} \simeq 0.529 \text{ \AA} .\tag{7.76}$$

Note that if the charge of the electron e^2 is decreased, the attraction force decreases and, correctly, the Bohr radius increases. The Bohr radius is the length scale of the hydrogen atom.

A natural energy scale E_0 is

$$E_0 = \frac{e^2}{a_0} = \frac{e^4 m}{\hbar^2} = \left(\frac{e^2}{\hbar c}\right)^2 mc^2 = \alpha^2 (mc^2) \quad (7.77)$$

where we see the appearance of the fine-structure constant α that, in cgs units, takes the form

$$\alpha \equiv \frac{e^2}{\hbar c} \simeq \frac{1}{137}. \quad (7.78)$$

We thus see that the natural energy scale of the hydrogen atom is about $\alpha^2 \simeq 1/18770$ smaller than the rest energy of the electron. This gives about $E_0 = 27.2\text{eV}$. In fact $-E_0/2 = -13.6\text{eV}$ is the bound state energy of the electron in the ground state of the hydrogen atom.

One curious way to approach the calculation of the ground state energy and ground state wavefunction is to factorize the Hamiltonian. One can show that

$$H = \gamma + \frac{1}{2m} \sum_{k=1}^3 \left(\hat{p}_k + i\beta \frac{\hat{x}_k}{r} \right) \left(\hat{p}_k - i\beta \frac{\hat{x}_k}{r} \right) \quad (7.79)$$

for suitable constants β and γ that you can calculate. The ground state $|\Psi_0\rangle$ is then the state for which

$$\left(\hat{p}_k - i\beta \frac{\hat{x}_k}{r} \right) |\Psi_0\rangle = 0. \quad (7.80)$$

The spectrum of the hydrogen atom is described in Figure 8. The energy levels are $E_{\nu\ell}$, where we used $\nu = 1, 2, \dots$, instead of n to label the various solutions for a given ℓ . This is because the label n is reserved for what is called the “principal quantum number”. The degeneracy of the system is such that multiplets with equal $n \equiv \nu + \ell$ have the same energy, as you can see in the figure. Thus, for example, $E_{2,0} = E_{1,1}$, which is to say that the first excited solution for $\ell = 0$ has the same energy as the lowest energy solution for $\ell = 1$. It is also important to note that for any fixed value of n the allowed values of ℓ are

$$\ell = 0, 1, \dots, n-1 \quad (7.81)$$

Finally, the energies are given by

$$E_{\nu\ell} = -\frac{e^2}{2a_0} \frac{1}{(\nu + \ell)^2}, \quad n \equiv \nu + \ell. \quad (7.82)$$

The large amount of degeneracy in this spectrum asks for an explanation. The hydrogen Hamiltonian has in fact some hidden symmetry. It has to do with the so-called Runge-Lenz vector. In the following we discuss the classical origin of this conserved vector quantity.

Imagine we have an energy functional

$$E = \frac{\mathbf{p}^2}{2m} + V(r) \quad (7.83)$$

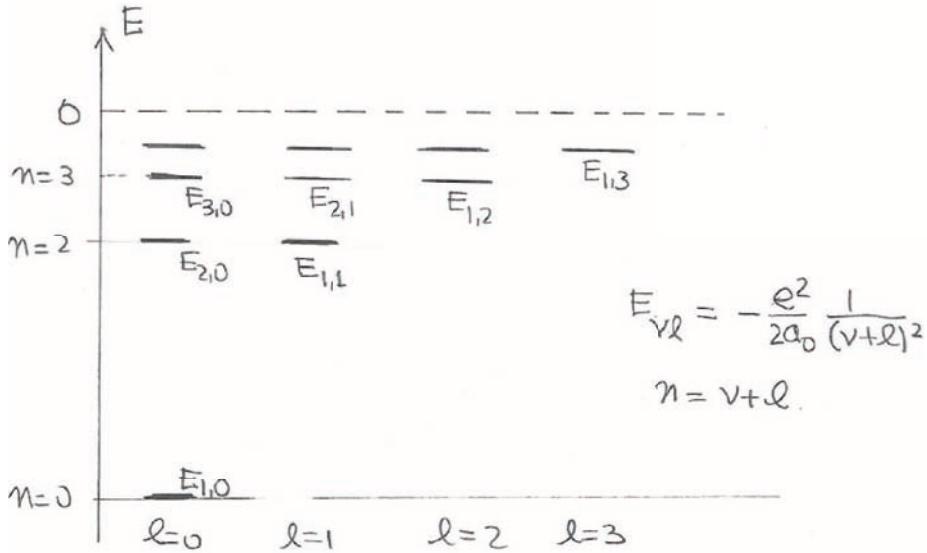


Figure 8: Spectrum of angular momentum multiplets for the hydrogen atom. Here $E_{\nu\ell}$ with $\nu = 1, 2, \dots$, denotes the energy of the ν -th solution for any fixed ℓ . States with equal values of $n \equiv \nu + \ell$ are degenerate. For any fixed n , the values of ℓ run from zero to $n - 1$. Correction: the $n = 0$ in the figure should be $n = 1$.

then the force on the particle moving in this potential is

$$\mathbf{F} = -\nabla V = -V'(r) \frac{\mathbf{r}}{r}, \quad (7.84)$$

where primes denote derivatives with respect to the argument. Newton's equation is

$$\frac{d\mathbf{p}}{dt} = -V'(r) \frac{\mathbf{r}}{r} \quad (7.85)$$

and it is simple to show (do it!) that in this central potential the angular momentum is conserved

$$\frac{d\mathbf{L}}{dt} = 0. \quad (7.86)$$

We now calculate (all classically) the time derivative of $\mathbf{p} \times \mathbf{L}$:

$$\begin{aligned} \frac{d}{dt}(\mathbf{p} \times \mathbf{L}) &= \frac{d\mathbf{p}}{dt} \times \mathbf{L} = -\frac{V'(r)}{r} \mathbf{r} \times (\mathbf{r} \times \mathbf{p}) \\ &= -\frac{mV'(r)}{r} \mathbf{r} \times (\mathbf{r} \times \dot{\mathbf{r}}) \\ &= -\frac{mV'(r)}{r} [\mathbf{r}(\mathbf{r} \cdot \dot{\mathbf{r}}) - \dot{\mathbf{r}} r^2]. \end{aligned} \quad (7.87)$$

We now note that

$$\mathbf{r} \cdot \dot{\mathbf{r}} = \frac{1}{2} \frac{d}{dt}(\mathbf{r} \cdot \mathbf{r}) = \frac{1}{2} \frac{d}{dt} r^2 = r \dot{r}. \quad (7.88)$$

Using this

$$\begin{aligned}\frac{d}{dt}(\mathbf{p} \times \mathbf{L}) &= -\frac{mV'(r)}{r} [\mathbf{r} r\dot{r} - \dot{\mathbf{r}} r^2] = mV'(r)r^2 \left[\frac{\dot{\mathbf{r}}}{r} - \frac{\mathbf{r} \dot{r}}{r^2} \right] \\ &= mV'(r)r^2 \frac{d}{dt} \left(\frac{\mathbf{r}}{r} \right)\end{aligned}\quad (7.89)$$

Because of the factor $V'(r)r^2$, the right-hand side fails to be a total time derivative. But if we focus on potentials for which this factor is a constant we will get a conservation law. So, assume

$$V'(r)r^2 = \gamma, \quad (7.90)$$

for some constant γ . Then

$$\frac{d}{dt}(\mathbf{p} \times \mathbf{L}) = m\gamma \frac{d}{dt} \left(\frac{\mathbf{r}}{r} \right) \rightarrow \frac{d}{dt} \left(\mathbf{p} \times \mathbf{L} - m\gamma \frac{\mathbf{r}}{r} \right) = 0 \quad (7.91)$$

We got a conservation law: that complicated vector inside the parenthesis is constant in time! Back to (7.90) we have

$$\frac{dV}{dr} = \frac{\gamma}{r^2} \rightarrow V(r) = -\frac{\gamma}{r} + c_0. \quad (7.92)$$

This is the most general potential for which we get a conservation law. For $c_0 = 0$ and $\gamma = e^2$ we have the hydrogen atom potential

$$V(r) = -\frac{e^2}{r}, \quad (7.93)$$

so we have

$$\frac{d}{dt} \left(\mathbf{p} \times \mathbf{L} - me^2 \frac{\mathbf{r}}{r} \right) = 0. \quad (7.94)$$

Factoring a constant we obtain the unit-free conserved **Runge-Lenz** vector \mathbf{R} :

$$\mathbf{R} \equiv \frac{1}{me^2} \mathbf{p} \times \mathbf{L} - \frac{\mathbf{r}}{r}, \quad \frac{d\mathbf{R}}{dt} = 0.$$

(7.95)

The conservation of the Runge-Lenz vector is a property of inverse squared central forces. The second vector in \mathbf{R} is simply minus the unit radial vector.

To understand the Runge-Lenz vector, we first examine its value for a circular orbit, as shown in figure 9. The vector \mathbf{L} is out of the page and $\mathbf{p} \times \mathbf{L}$ points radially outward. The vector \mathbf{R} is thus a competition between the outward radial first term and the inner radial second term. If these two terms would not cancel, the result would be a radial vector (outwards or inwards) but in any case, not conserved, as it rotates with the particle. Happily, the two terms cancel. Indeed for a circular orbit

$$m \frac{v^2}{r} = \frac{e^2}{r^2} \rightarrow \frac{m^2 v^2 r}{me^2} = 1 \rightarrow \frac{(mv)(mvr)}{me^2} = 1 \rightarrow \frac{pL}{me^2} = 1, \quad (7.96)$$

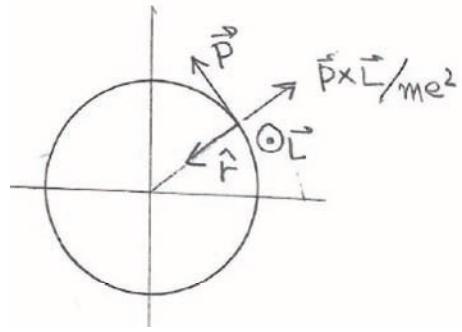


Figure 9: The Runge-Lenz vector vanishes for a circular orbit.

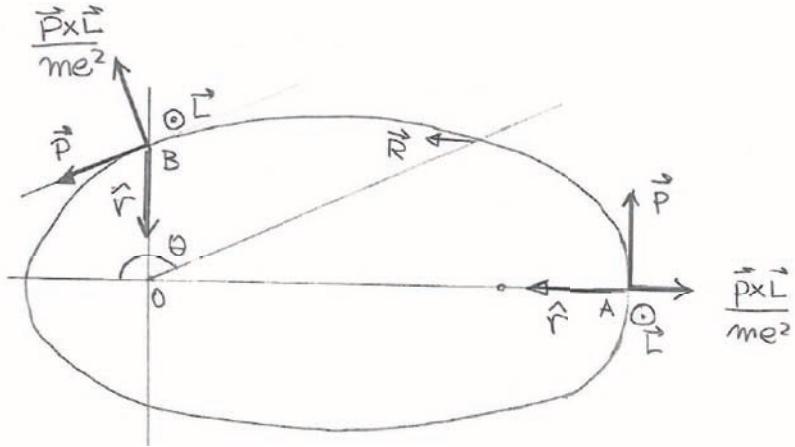


Figure 10: In an elliptic orbit the Runge-Lenz vector is a vector along the major axis of the ellipse and points in the direction from the aphelion to the focus.

which is the statement that the first vector in \mathbf{R} , for a circular orbit, is of unit length and being outward directed cancels with the second term. The Runge-Lenz vector indeed vanishes for a circular orbit.

We now argue that for an elliptical orbit the Runge-Lenz vector is not zero. Consider figure 10. At the aphelion (point furthest away from the focal center), denoted as point *A* we have the first term in \mathbf{R} point outwards and the second term point inwards. Thus, if \mathbf{R} does not vanish it must be a vector along the line joining the focus and the aphelion, a horizontal vector on the figure. Now consider point *B* right above the focal center of the orbit. At this point \mathbf{p} is no longer perpendicular to the radial vector and therefore $\mathbf{p} \times \mathbf{L}$ is no longer radial. As you can see, it points slightly to the left. It follows that \mathbf{R} points to the left side of the figure. \mathbf{R} is a vector along the major axis of the ellipse and points in the direction from the aphelion to the focus.

To see more quantitatively the role of \mathbf{R} we dot its definition with the radial vector \mathbf{r} :

$$\mathbf{r} \cdot \mathbf{R} = \frac{1}{me^2} \mathbf{r} \cdot (\mathbf{p} \times \mathbf{L}) - r \quad (7.97)$$

referring to the figure, with the angle θ as defined there and $R \equiv |\mathbf{R}|$, we get

$$rR \cos \theta = \frac{1}{me^2} \mathbf{L} \cdot (\mathbf{r} \times \mathbf{p}) - r = \frac{1}{me^2} L^2 - r. \quad (7.98)$$

Collecting terms proportional to r :

$$r(1 + R \cos \theta) = \frac{L^2}{me^2} \rightarrow \boxed{\frac{1}{r} = \frac{me^2}{L^2}(1 + R \cos \theta)}, \quad (7.99)$$

We identify the magnitude R of the Runge-Lenz vector with the eccentricity of the orbit! Indeed if $R = 0$ the orbit is circular: r does not depend on θ .

This whole analysis has been classical. Quantum mechanically we will need to change some things a bit. The definition of \mathbf{R} only has to be changed to guarantee that \mathbf{R} is a hermitian (vector) operator. As you will verify the hermitization gives

$$\mathbf{R} \equiv \frac{1}{2me^2} (\mathbf{p} \times \mathbf{L} - \mathbf{L} \times \mathbf{p}) - \frac{\mathbf{r}}{r} \quad (7.100)$$

The quantum mechanical conservation of \mathbf{R} is the statement that it commutes with the hydrogen Hamiltonian

$$[\mathbf{R}, H] = 0. \quad (7.101)$$

You will verify this; it is the analog of our classical calculation that showed that the time-derivative of \mathbf{R} is zero. Moreover, the length-squared of the vector is also of interest. You will show that

$$\mathbf{R}^2 = 1 + \frac{2}{me^4} H(\mathbf{L}^2 + \hbar^2). \quad (7.102)$$

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

PARTICLE NATURE OF LIGHT AND WAVE NATURE OF MATTER

B. Zwiebach
February 16, 2016

Contents

1 Photoelectric Effect	1
2 Compton Scattering	3
3 Matter Waves	6

1 Photoelectric Effect

The photoelectric effect was first observed by Heinrich Hertz in 1887. When polished metal plates are irradiated, they may emit electrons, then called “photo-electrons”. The emitted electrons thus produce a *photoelectric current*. The key observations were:

- There is a threshold frequency ν_0 . Only for frequencies $\nu > \nu_0$ is there a photoelectric current. The frequency ν_0 depends on the metal and the configuration of the atoms at the surface. It is also affected by inhomogeneities.
- The magnitude of the photoelectric current is proportional to the intensity of the light source.
- Energy of the photoelectrons is *independent* of the intensity of the light source.

A natural explanation for the features in this effect didn't come until 1905, when Einstein explained the above features by postulating that the energy in light is carried by discrete quanta (later called photons) with energy $h\nu$. Here h is Planck's constant, the constant used by Planck to produce a fit for the blackbody energy as a function of frequency.

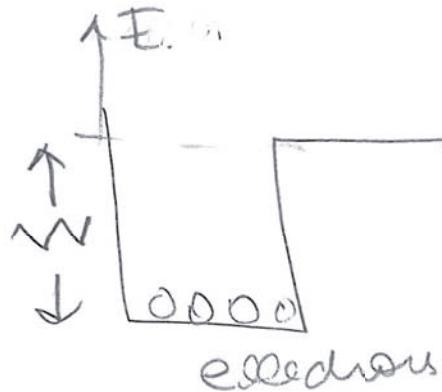


Figure 1: Electrons in a metal are bound. If the photon energy is greater than the work function W an electron may be ejected.

A given material has a characteristic energy W , called the *work function*, which is the minimum energy required to eject an electron. This is not easily calculated because it is the result of an

interaction of many electrons with the background of atoms. It is easily measured, however. When the surface of the material is irradiated, electrons in the material absorb the energy of the incoming photons. If the energy imparted on an electron by the absorption of a single photon is greater than the work function W , then the electron is ejected with kinetic energy E_{e^-} equal to the difference of the photon energy and the work function:

$$E_{e^-} = \frac{1}{2}mv^2 = h\nu - W = E_\gamma - W. \quad (1.1)$$

This equation, written by Einstein explains the experimental features noted above, once we assume that the quanta act on individual electrons to eject them. The threshold frequency is defined by

$$h\nu_0 = W, \quad (1.2)$$

as it leads to a photoelectron with zero energy. For $\nu > \nu_0$ the electrons will be ejected. Increasing the intensity of the light source increases the rate that photons arrive, which will increase the magnitude of the current, but will not change the energy of the photoelectrons because it does not change the energy of each incoming quanta.

Equation (1.2) allowed Einstein to make a prediction: The kinetic energy of the photo-electrons increases linearly with the frequency of light. Einstein's prediction was confirmed experimentally by Millikan (1915) who measured carefully the photoelectron energies and confirmed their linear dependence on the energy. Millikan's careful work allowed him to determine the value of Planck's constant \hbar to better than 1% accuracy! Still, skepticism remained and physicists were not yet convinced about the particle nature of these light quanta.

Example: Consider UV light with wavelength $\lambda = 290\text{nm}$ incident on a metal with work function $W = 4.05\text{eV}$. What is the energy of the photo-electron and what is its speed?

Solution: It is useful to solve these problems without having to look up constants. For this try recalling this useful relation

$$\hbar c = 197.33 \text{ MeV.fm}, \quad \hbar \equiv \frac{h}{2\pi}, \quad (1.3)$$

where $\text{MeV} = 10^6\text{eV}$ and $\text{fm} = 10^{-15}\text{m}$. Let us use this to compute the photon energy. In this case,

$$E_\gamma = h\nu = 2\pi\hbar\frac{c}{\lambda} = \frac{2\pi \cdot 197.33 \text{ MeV.fm}}{290 \times 10^{-9}\text{m}} = \frac{2\pi \cdot 197.33}{290} \text{ eV} \approx 4.28 \text{ eV}, \quad (1.4)$$

and thus

$$E_{e^-} = E_\gamma - W = 0.23 \text{ eV}. \quad (1.5)$$

To compute the energy we set

$$0.23 \text{ eV} = \frac{1}{2}m_e v^2 = \frac{1}{2}(m_e c^2) \left(\frac{v}{c}\right)^2 \quad (1.6)$$

Recalling that $m_e c^2 \simeq 511,000\text{eV}$ one finds

$$\frac{0.46}{511000} = \left(\frac{v}{c}\right)^2 \rightarrow \frac{v}{c} = 0.0009488. \quad (1.7)$$

With and $c = 300,000 \text{ Km/s}$ we finally get $v \simeq 284.4 \text{ Km/s}$.

This is a good point to consider units, in particular the units of h . We can ask: Is there a physical quantity that has the units of h . The answer is yes, as we will see now. From the equation $E = h\nu$, we have

$$[h] = \left[\frac{E}{\nu}\right] = \frac{ML^2/T^2}{1/T} = L \cdot M \frac{L}{T}, \quad (1.8)$$

where $[\cdot]$ gives the units of a quantity, and M, L, T are units of mass, length, and time, respectively. We have written the right-most expression as a product of units of length and momentum. Therefore

$$[h] = [\mathbf{r} \times \mathbf{p}] = [\mathbf{L}]. \quad (1.9)$$

We see that h has units of angular momentum! Indeed for a spin one-half particle, the magnitude of the spin angular momentum is $\frac{1}{2}\hbar$.

With $[h] = [r][p]$ we also see that one has a canonical way to associate a length to any particle of a given mass m . Indeed, using the speed of light, we can construct the momentum $p = mc$, and then the length ℓ is obtained from the ratio h/p . This actually is the **Compton wavelength** λ_C of a particle:

$$\lambda_C = \frac{h}{mc} \quad (1.10)$$

then has units of length; this is called the *Compton wavelength* of a particle of mass m . Note that this length is independent of the velocity of the particle. The de Broglie wavelength of the particle uses the true momentum of the particle, not mc ! Thus, Compton and de Broglie wavelengths should not be confused!

It is possible to get some physical intuition for the Compton wavelength λ_C of a particle. We claim that λ_C is the wavelength of a photon whose energy is equal to the rest energy of the particle. Indeed we would have

$$mc^2 = h\nu = h\frac{c}{\lambda} \rightarrow \lambda = \frac{h}{mc}, \quad (1.11)$$

confirming the claim. Suppose you are trying to localize a point particle of mass m . If you use light, the possible accuracy in the position of the particle is roughly the wavelength of the light. Once we use light with $\lambda < \lambda_C$ the photons carry more energy than the rest energy of the particle. It is possible then that the energy of the photons go into creating more particles of mass m , making it difficult, if not impossible to localize the particle. The Compton wavelength is the length scale at which we need *relativistic quantum field theory* to take into account the possible processes of particle creation and annihilation.

Let us calculate the Compton wavelength of the electron:

$$\lambda_C(e) = \frac{h}{m_e c} = \frac{2\pi\hbar c}{m_e c^2} = \frac{2\pi \cdot 197.33 \text{ MeV.fm}}{0.511 \text{ MeV}} = 2426 \text{ fm} = 2.426 \text{ pm}. \quad (1.12)$$

This length is about 20 times smaller than the Bohr radius (53 pm.) and about two-thousand times the size of a proton (1 fm.). The Compton wavelength of the electron appears in the formula for the change of photon wavelength in the process called Compton scattering.

2 Compton Scattering

Originally Einstein did not make clear that the light quantum meant a particle of light. In 1916, however, he posited that the quantum would carry momentum as well as energy, making the case for a particle much clearer. In relativity, the energy, momentum, and rest mass of a particle are related by

$$E^2 - p^2 c^2 = m^2 c^4. \quad (2.13)$$

(Compare this with the classical equation $E = p^2/2m$.) Of course, one can also express the energy and momentum of the particle in terms of the velocity:

$$E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad \mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (2.14)$$

You should use these expressions to confirm that (2.13) holds ($|\mathbf{p}| = p$). A particle that moves with the speed of light, like the photon, must have zero rest mass, otherwise its energy and momentum would be infinite due to the vanishing denominators. With the rest mass set to zero, equation (2.13) gives the relation between the photon energy E_γ and the photon momentum p_γ :

$$E_\gamma = p_\gamma c. \quad (2.15)$$

Then, using $\lambda\nu = c$, we reach

$$p_\gamma = \frac{E_\gamma}{c} = \frac{h\nu}{c} = \frac{h}{\lambda}. \quad (2.16)$$

We will see this relation again later when we discuss matter waves.

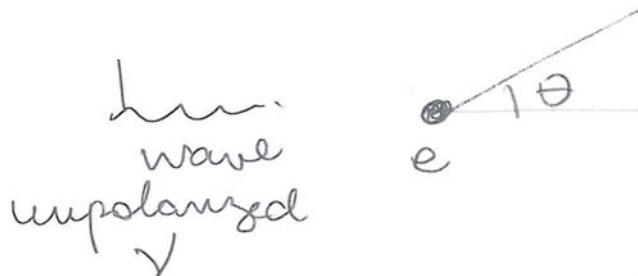


Figure 2: Unpolarized light incident on an electron scatters into an angle θ . Classically, this is described by Thomson scattering. The light does not change frequency during this process.

Compton carried out experiments (1923–1924) scattering X-rays off a carbon target. X-rays correspond to photon energies in the range from 100 eV to 100 KeV. The goal was scattering X-ray photons off free electrons, and with some qualification, the electrons in the atoms behave this way.

The classical counterpart of the Compton experiment is the scattering of electromagnetic waves off free electrons, called *Thompson scattering*. Here an electromagnetic wave is incident on an electron. The electric field of the wave shakes the electron which oscillates with the frequency of the incoming field. The electron oscillation produces a radiated field, of the same frequency as that of the incoming radiation. In classical Thomson scattering the differential scattering cross section is given by

$$\frac{d\sigma}{d\Omega} = \left(\frac{e^2}{mc^2} \right)^2 \frac{1}{2} (1 + \cos^2 \theta), \quad (2.17)$$

where θ is the angle between the incident and scattered wave, with the radiated energy at the same frequency as the incoming light. This is shown in Figure 2. The cross-section has units of length-squared, or area, as it should. It represents the area that would extract from the incoming plane wave the amount of energy that is scattered by the electron. Indeed the quantity $e^2/(mc^2)$ is called the classical electron radius and it is about 2.8 fm! not much bigger than a proton!

If we treat the light as photons, the elementary process going on is a collision between two particles; an incoming photon and a roughly stationary electron. Two facts can be quickly demonstrated:

- The photon cannot be absorbed by the electron. It is inconsistent with energy and momentum conservation (exercise)
- The photon must lose some energy and thus the final photon wavelength λ_f must be larger than the initial photon wavelength λ_i . This is clear in the laboratory frame, where the initially stationary electron must recoil and thus acquire some kinetic energy.

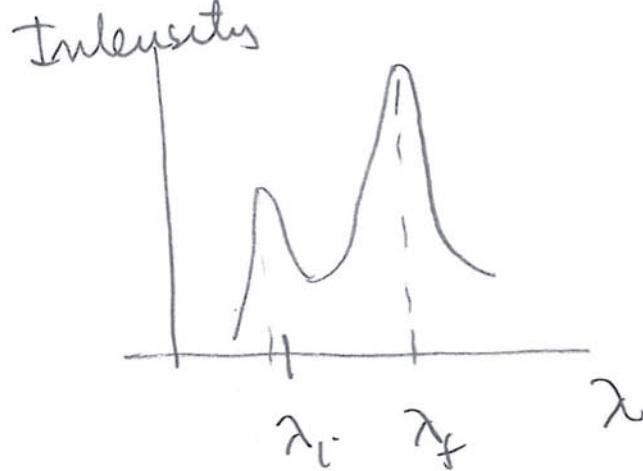


Figure 3: The results of Compton's scattering experiment. The incident photon wavelength is λ_i , and the scattered photon wavelength is $\lambda_f \simeq \lambda_i + \ell_C$, corresponding to $\theta = 90^\circ$.

Indeed, Compton's observations did not agree with the predictions of Thompson scattering: the X-rays changed frequency after scattering. A calculation using energy and momentum conservation shows that the change of wavelength is correlated with the angle between the scattered photon and the original photon:

$$\lambda_f = \lambda_i + \frac{h}{m_e c} (1 - \cos \theta) = \lambda_i + \ell_C (1 - \cos \theta). \quad (2.18)$$

Note that appearance of the Compton wavelength of the electron, the particle the photon scatters off from. The maximum energy loss for the photon occurs at $\theta = \pi$, where

$$\lambda_f(\theta = 180^\circ) = \lambda_i + 2\lambda_C. \quad (2.19)$$

The maximum possible change in wavelength is $2\lambda_C$. For $\theta = \frac{\pi}{2}$ the change of wavelength is exactly ℓ_C

$$\lambda_f(\theta = 90^\circ) = \lambda_i + \lambda_C. \quad (2.20)$$

Compton's experiment used molybdenum X-rays with energy and wavelength

$$E_\gamma \approx 17.5 \text{ keV}, \quad \lambda_i = 0.0709 \text{ nm}, \quad (2.21)$$

incident on a carbon target. Placing the detector at an angle $\theta = 90^\circ$ the plot of the intensity (or number of photons scattered) as a function of wavelength is shown in Figure 2. One finds a peak for $\lambda_f = 0.0731 \text{ nm}$, but also a second peak at the original wavelength $\lambda_i = 0.0709 \text{ nm}$.

The peak at λ_f is the expected one: $\lambda_f - \lambda_i \simeq 2.2 \text{ pm}$, which is about the Compton wavelength of 2.4 pm. Given that the photons have energies of 17 KeV and the bound state energies of carbon

are about 300 eV, the expected peak represents instances where the atom is ionized by the collision and it is a fine approximation to consider the ejected electrons. The peak at λ_i represents a process in which an electron receives some momentum from the photon but still remains bound. This is not very unlikely: the typical momentum of a bound electron is actually comparable to the momentum of the photon. In this case the photon scatters at 90° and the recoil momentum is carried by the whole atom. The relevant Compton wavelength is therefore that of the atom. Since the mass of the carbon atom is several thousands of times larger than the mass of the electron, the Compton wavelength of the atom is much smaller than the electron Compton wavelength and there should be no detectable change in the wavelength of the photon.¹

3 Matter Waves

As we have seen, light behaves as both a particle and a wave. This kind of behavior is usually said to be a **duality**: the complete reality of the object is captured using *both* the wave and particle features of the object. The photon is a particle of energy E_γ , but has frequency ν which is a wave attribute, with $E = h\nu$. It is a particle with momentum p_γ but it also has a wavelength λ , a wave attribute, given by (2.16)

$$\lambda = \frac{h}{p_\gamma}. \quad (3.22)$$

In 1924, Louis de Broglie proposed that the wave/particle duality of the photon was universal, and thus valid for matter particles too. In this way he conjectured the *wave nature of matter*. Inspired by (3.22) de Broglie postulated that associated to a matter particle with momentum p there is a plane wave of wavelength λ given by

$$\lambda = \frac{h}{p}. \quad (3.23)$$

This is a fully quantum property: if $h \rightarrow 0$, then $\lambda \rightarrow 0$, and the particles have no wave properties. And exciting consequence of this is that matter particles can diffract or interfere! In the famous Davisson-Germer experiment (1927) electrons are strike a metal surface and one finds that at certain angles there are peaks in the intensity of the scattered electrons. The peaks showed the effect of constructive interference from scattering off the lattice of atoms in the metal, demonstrating the wave nature of the electrons. One can also do two-slit interference with electrons, and the experiment can be done shooting one electron at a time. A recent experiment [arXiv:1310.8343] by Eibenberger *et.al* reports interference using molecules with 810 atoms and mass exceeding 10 000 amu (that's 20 million times the mass of the electron!).

The de Broglie wavelength can be calculated to estimate if quantum effects are important. Consider for this purpose a particle of mass m and momentum p incident upon an object of size x , as illustrated in Figure 3. Let $\lambda = h/p$ denote the de Broglie wavelength of the particle. The wave nature of the particle is not important if λ is much smaller than x . Thus, the “classical approximation,” in which wave effects are negligible, requires

$$\text{Wave effects negligible: } \frac{\lambda}{x} \ll 1. \quad (3.24)$$

Using $\lambda = h/p$, this yields

$$\text{Wave effects negligible: } x p \gg h, \quad (3.25)$$

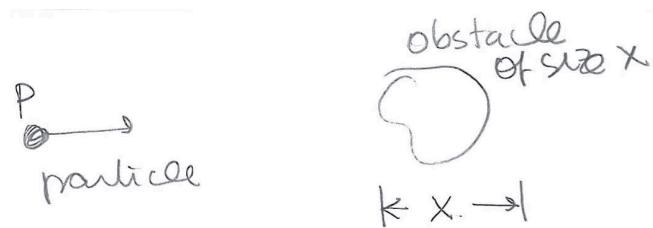


Figure 4: A particle of momentum p incident on an obstacle of size x .

a relation in which both sides have units of angular momentum.

Classical behavior is a subtle limit of quantum mechanics: a classical electromagnetic field requires a large number of photons. Any state with an exact, fixed number of photons, even if large, is not classical, however. Classical electromagnetic states are so-called coherent states, in which the number of photons fluctuates.

Andrew Turner transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

¹Thanks to V. Vuletic for a clarification of this point.

MIT OpenCourseWare
<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

MULTIPARTICLE STATES AND TENSOR PRODUCTS

B. Zwiebach

December 14, 2013

Contents

1	Introduction to the Tensor Product	1
2	Entangled States	8
3	Bell basis states	9
4	Quantum Teleportation	13
5	EPR and Bell Inequalities	16

1 Introduction to the Tensor Product

In this section, we develop the tools needed to describe a system that contains more than one particle. Most of the required ideas appear when we consider systems with two particles. We will assume the particles are distinguishable; for indistinguishable particles quantum mechanics imposes some additional constraints on the allowed set of states. We will study those constraints later in the course (or in 8.06!) The tools we are about to develop will be needed to understand addition of angular momenta. In that problem one is adding the angular momenta of the two or more particles in the system.

Consider then two particles. Below is a description of the quantum mechanics and family of operators associated with each particle:

- Particle 1: its quantum mechanics is described by a complex vector space V . It has associated operators T_1, T_2, \dots
- Particle 2: its quantum mechanics is described by a complex vector space W . It has associated operators S_1, S_2, \dots

This list of operators for each particle may include some or many of the operators you are already familiar with: position, momentum, spin, Hamiltonians, projectors, etc.

Once we have two particles, the two of them together form our system. We are after the description of quantum states of this two-particle system. On first thought, we may think that any state of this system should be described by giving the state $v \in V$ of the first particle and the state $w \in W$ of the second particle. This information could be represented by the ordered list (v, w) where the first item

is the state of the first particle and the second item the state of the second particle. This *is* a state of the two-particle system, but it is far from being the general state of the two-particle system. It misses remarkable new possibilities, as we shall soon see.

We thus introduce a new notation. Instead of representing the state of the two-particle system with particle one in v and particle two in w as (v, w) , we will represent it as $v \otimes w$. This element $v \otimes w$ will be viewed as a vector in a new vector space $V \otimes W$ that will carry the description of the quantum states of the system of two particles. This \otimes operation is called the “tensor product.” In this case we have two vector spaces over \mathbb{C} and the tensor product $V \otimes W$ is a new complex vector space:

$$v \otimes w \in V \otimes W \quad \text{when } v \in V, w \in W. \quad (1.1)$$

In $v \otimes w$ there is no multiplication to be carried out, we are just placing one vector to the left of \otimes and another to the right of \otimes .

We have only described some elements of $V \otimes W$, not quite given its definition yet.¹ We now explain two physically motivated rules that define the tensor product completely.

1. If the vector representing the state of the first particle is scaled by a complex number this is equivalent to scaling the state of the two particles. The same for the second particle. So we declare

$$(av) \otimes w = v \otimes (aw) = a(v \otimes w), \quad a \in \mathbb{C}. \quad (1.2)$$

2. If the state of the first particle is a superposition of two states, the state of the two-particle system is also a superposition. We thus demand distributive properties for the tensor product:

$$\begin{aligned} (v_1 + v_2) \otimes w &= v_1 \otimes w + v_2 \otimes w, \\ v \otimes (w_1 + w_2) &= v \otimes w_1 + v \otimes w_2. \end{aligned} \quad (1.3)$$

The tensor product $V \otimes W$ is thus defined to be the vector space whose elements are (complex) linear combinations of elements of the form $v \otimes w$, with $v \in V, w \in W$, with the above rules for manipulation. The tensor product $V \otimes W$ is the complex vector space of states of the two-particle system!

Comments

1. The vector $0 \in V \otimes W$ is equal to $0 \otimes w$ or $v \otimes 0$. Indeed, by the first property above, with $a = 0$, we have $av = 0$ (rhs a vector) and $0 \otimes w = 0(0 \otimes w) = 0$
2. Let $v_1, v_2 \in V$ and $w_1, w_2 \in W$. A vector in $V \otimes W$ constructed by superposition is

$$\alpha_1(v_1 \otimes w_1) + \alpha_2(v_2 \otimes w_2) \in V \otimes W \quad (1.4)$$

¹If we just left it like this, we would have defined the direct product of vector spaces.

This shows clearly that a general state of the two-particle system cannot be described by stating the state of the first particle and the state of the second particle. The above superpositions give rise to entangled states. An entangled state of the two particles is one that, roughly, cannot be disentangled into separate states of each of the particles. We will make this precise soon.

If (e_1, \dots, e_n) is a basis of V and (f_1, \dots, f_m) is a basis of W , then the set of elements $e_i \otimes f_j$ where $i = 1, \dots, n$ and $j = 1, \dots, m$ forms a basis for $V \otimes W$. It is simple to see these span the space since for any $v \otimes w$ we have $v = \sum_i v_i e_i$ and $w = \sum_j w_j f_j$ so that

$$v \otimes w = (\sum_i v_i e_i) \otimes (\sum_j w_j f_j) = \sum_{i,j} v_i w_j e_i \otimes f_j. \quad (1.5)$$

Given this, we see that the basis also spans linear superpositions of elements of the form $v \otimes w$, thus general elements of $V \otimes W$. With $n \cdot m$ basis vectors, the dimensionality of $V \otimes W$ is equal to the *product* of the dimensionalities of V and W :

$$\dim(V \otimes W) = \dim(V) \times \dim(W). \quad (1.6)$$

Dimensions are multiplied (not added) in a tensor product.

How do we construct operators that act in the vector space $V \otimes W$? Let T be an operator in V and S be an operator in W . In other words, $T \in \mathcal{L}(V)$ and $S \in \mathcal{L}(W)$. We can then construct an operator $T \otimes S$

$$T \otimes S \in \mathcal{L}(V \otimes W) \quad (1.7)$$

defined to act as follows:

$$T \otimes S (v \otimes w) \equiv T v \otimes S w. \quad (1.8)$$

This is the only ‘natural’ option: we let T act on the vector it knows how to act, and S act on the vector it knows how to act.

Suppose that we want the operator $T \in \mathcal{L}(V)$ that acts on the first particle to act on the tensor product $V \otimes W$, even though we have not supplied an operator S to act on the W part. For this we upgrade the operator from one that acts on a single vector space to one, given by $T \otimes 1$, that acts on the tensor product:

$$T \in \mathcal{L}(V) \rightarrow T \otimes 1 \in \mathcal{L}(V \otimes W), \quad T \otimes 1 (v \otimes w) \equiv T v \otimes w. \quad (1.9)$$

Similarly, an operator S belonging to $\mathcal{L}(W)$ is upgraded to $1 \otimes S$ to act on the tensor product. A basic result is that upgraded operators of the first particle **commute** with upgraded operators of the second particle. Indeed,

$$\begin{aligned} (T \otimes 1) \cdot (1 \otimes S) (v \otimes w) &= (T \otimes 1)(v \otimes S w) = T v \otimes S w \\ (1 \otimes S) \cdot (T \otimes 1) (v \otimes w) &= (1 \otimes S)(T v \otimes w) = T v \otimes S w. \end{aligned} \quad (1.10)$$

and therefore

$$[T \otimes \mathbf{1}, \mathbf{1} \otimes S] = 0. \quad (1.11)$$

Given a system of two particles we can construct a simple total Hamiltonian H_T (describing no interactions) by upgrading each of the Hamiltonians H_1 and H_2 and adding them:

$$H_T \equiv H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2 \quad (1.12)$$

Exercise. Convince yourself that

$$\exp\left(-\frac{iH_T t}{\hbar}\right) = \exp\left(-\frac{iH_1 t}{\hbar}\right) \otimes \exp\left(-\frac{iH_2 t}{\hbar}\right) \quad (1.13)$$

We turn now to a famous example at the basis of adding angular momenta.

Example 1: We have two spin-1/2 particles, and describe the first's state space V_1 with basis states $|+\rangle_1$ and $|-\rangle_1$ and the second's state space V_2 with basis states $|+\rangle_2$ and $|-\rangle_2$. The tensor product $V_1 \otimes V_2$ has four basis vectors:

$$|+\rangle_1 \otimes |+\rangle_2; \quad |+\rangle_1 \otimes |-\rangle_2; \quad |-\rangle_1 \otimes |+\rangle_2; \quad |-\rangle_1 \otimes |-\rangle_2 \quad (1.14)$$

If we follow the convention that the first ket corresponds to particle one and the second ket corresponds to particle two, the notation is simpler. The most general state of the two-particle system is a linear superposition of the four basis states:

$$|\Psi\rangle = \alpha_1|+\rangle_1 \otimes |+\rangle_2 + \alpha_2|+\rangle_1 \otimes |-\rangle_2 + \alpha_3|-\rangle_1 \otimes |+\rangle_2 + \alpha_4|-\rangle_1 \otimes |-\rangle_2. \quad (1.15)$$

Example 2: We now want to act on this state with the *total z-component* of angular momentum. Naively, this would be the sum of the z -components of each individual particle. However, we know better at this point - summing the two angular momenta really means constructing a new operator in the tensor product vector space:

$$S_z^T = S_z^{(1)} \otimes \mathbf{1} + \mathbf{1} \otimes S_z^{(2)}. \quad (1.16)$$

Performing the calculation in two parts,

$$\begin{aligned} (S_z^{(1)} \otimes \mathbf{1})|\Psi\rangle &= \alpha_1 S_z |+\rangle_1 \otimes |+\rangle_2 + \alpha_2 S_z |+\rangle_1 \otimes |-\rangle_2 + \alpha_3 S_z |-\rangle_1 \otimes |+\rangle_2 + \alpha_4 S_z |-\rangle_1 \otimes |-\rangle_2 \\ &= \frac{\hbar}{2} (\alpha_1 |+\rangle_1 \otimes |+\rangle_2 + \alpha_2 |+\rangle_1 \otimes |-\rangle_2 - \alpha_3 |-\rangle_1 \otimes |+\rangle_2 - \alpha_4 |-\rangle_1 \otimes |-\rangle_2) \\ (1 \otimes S_z^{(2)})|\Psi\rangle &= \alpha_1 |+\rangle_1 \otimes S_z |+\rangle_2 + \alpha_2 |+\rangle_1 \otimes S_z |-\rangle_2 + \alpha_3 |-\rangle_1 \otimes S_z |+\rangle_2 + \alpha_4 |-\rangle_1 \otimes S_z |-\rangle_2 \\ &= \frac{\hbar}{2} (\alpha_1 |+\rangle_1 \otimes |+\rangle_2 - \alpha_2 |+\rangle_1 \otimes |-\rangle_2 + \alpha_3 |-\rangle_1 \otimes |+\rangle_2 - \alpha_4 |-\rangle_1 \otimes |-\rangle_2) \end{aligned} \quad (1.17)$$

Adding these together, we have:

$$S_z^T |\Psi\rangle = \hbar (\alpha_1 |+\rangle_1 \otimes |+\rangle_2 - \alpha_4 |-\rangle_1 \otimes |-\rangle_2) \quad (1.18)$$

One can derive this result quickly by noting that since $S_z^{(1)}$ is diagonal in the first basis and $S_z^{(2)}$ is diagonal in the second basis, the total S_z is diagonal in the tensor space basis and its eigenvalue acting on a tensor state is the sum of the S_z eigenvalues for particle one and particle two. Thus,

$$\begin{aligned} S_z^T |+\rangle \otimes |+\rangle &= \left(\frac{\hbar}{2} + \frac{\hbar}{2}\right) |+\rangle \otimes |+\rangle = \hbar |+\rangle \otimes |+\rangle \\ S_z^T |+\rangle \otimes |-\rangle &= \left(\frac{\hbar}{2} - \frac{\hbar}{2}\right) |+\rangle \otimes |+\rangle = 0 \\ S_z^T |-\rangle \otimes |+\rangle &= \left(-\frac{\hbar}{2} + \frac{\hbar}{2}\right) |+\rangle \otimes |+\rangle = 0 \\ S_z^T |-\rangle \otimes |-\rangle &= \left(-\frac{\hbar}{2} - \frac{\hbar}{2}\right) |-\rangle \otimes |-\rangle = -\hbar |-\rangle \otimes |-\rangle \end{aligned} \quad (1.19)$$

The result in (1.18) follows quickly from the four relations above. Suppose we are only interested in states that have zero S_z^T . This requires

$$\alpha_1 = \alpha_4 = 0 \rightarrow |\Psi\rangle = \alpha_2 |+\rangle \otimes |-\rangle + \alpha_3 |-\rangle \otimes |+\rangle \quad (1.20)$$

Example 3: Calculate the total x -component S_x^T of spin angular momentum on the above states with zero S_z^T . Recalling that

$$S_x |+\rangle = \frac{\hbar}{2} |-\rangle, \quad S_x |-\rangle = \frac{\hbar}{2} |+\rangle \quad (1.21)$$

and writing

$$S_x^T = S_x \otimes 1 + 1 \otimes S_x \quad (1.22)$$

the calculation proceeds as follows:

$$\begin{aligned} S_x^T |+\rangle \otimes |-\rangle &= S_x |+\rangle \otimes |-\rangle + |+\rangle \otimes S_x |-\rangle = \frac{\hbar}{2} (|-\rangle \otimes |-\rangle + |+\rangle \otimes |+\rangle) \\ S_x^T |-\rangle \otimes |+\rangle &= S_x |-\rangle \otimes |+\rangle + |-\rangle \otimes S_x |+\rangle = \frac{\hbar}{2} (|+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle) \end{aligned} \quad (1.23)$$

Therefore

$$\begin{aligned} S_x^T |\Psi\rangle &= \alpha_2 \frac{\hbar}{2} (|-\rangle \otimes |-\rangle + |+\rangle \otimes |+\rangle) + \alpha_3 \frac{\hbar}{2} (|+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle) \\ &= \frac{\hbar}{2} (\alpha_2 + \alpha_3) (|+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle) \end{aligned} \quad (1.24)$$

If we demand that S_x^T also be zero on the state we now find $\alpha_2 = -\alpha_3$. Thus, the following state has zero S_x^T, S_z^T :

$$|\Psi\rangle = \alpha (|+\rangle \otimes |-\rangle - |-\rangle \otimes |+\rangle). \quad (1.25)$$

Exercise: Verify that $S_y^T |\Psi\rangle = 0$. Thus we say that the state has total spin angular momentum zero.

We now consider the definition of an **inner product** in $V \otimes W$. To do this we simply give state how the most general inner product is computed using a basis $\{e_i \otimes f_j\}$ for the tensor product, with $\{e_i\}$ and $\{f_i\}$ *orthonormal* bases for V and W . We begin by declaring that

$$\langle e_i \otimes f_j, e_p \otimes f_q \rangle \equiv \delta_{ip} \delta_{jq}. \quad (1.26)$$

This makes the basis $\{e_i \otimes f_j\}$ orthonormal. In addition, we must declare that with vectors $X, Y, Z \in V \otimes W$ and a complex constant a the following axioms hold:

$$\begin{aligned} \langle X + Y, Z \rangle &= \langle X, Z \rangle + \langle Y, Z \rangle, \\ \langle X, Y + Z \rangle &= \langle X, Y \rangle + \langle X, Z \rangle, \\ \langle X, aY \rangle &= a \langle X, Y \rangle \\ \langle aX, Y \rangle &= a^* \langle X, Y \rangle \end{aligned} \quad (1.27)$$

This is a complete definition of the inner product in the tensor space: we can compute the inner product of any two vectors in $V \otimes W$ using the chosen basis and the above distributive rules. Indeed, using these properties we can show that

$$\langle v \otimes w, \tilde{v} \otimes \tilde{w} \rangle = \langle v, \tilde{v} \rangle \langle w, \tilde{w} \rangle, \quad (1.28)$$

where the inner products on the right-hand side are those in V and in W , making it clear that the inner product in $V \otimes W$ arises from the inner products in V and W . To prove this relation we begin by writing

$$\begin{aligned} v &= \sum_i v_i e_i, & w &= \sum_j w_j f_j, \\ \tilde{v} &= \sum_p \tilde{v}_p e_p, & \tilde{w} &= \sum_q \tilde{w}_q f_q. \end{aligned} \quad (1.29)$$

Since the basis vectors in V and W are orthonormal we find that

$$\langle v, \tilde{v} \rangle = \sum_i v_i^* \tilde{v}_i, \quad \langle w, \tilde{w} \rangle = \sum_j w_j^* \tilde{w}_j. \quad (1.30)$$

Now evaluating the left-hand side of (1.28)

$$\begin{aligned} \langle v \otimes w, \tilde{v} \otimes \tilde{w} \rangle &= \left\langle \sum_i v_i e_i \otimes \sum_j w_j f_j, \sum_p \tilde{v}_p e_p \otimes \sum_q \tilde{w}_q f_q \right\rangle \\ &= \sum_{i,j,p,q} \left\langle v_i w_j e_i \otimes f_j, \tilde{v}_p \tilde{w}_q e_p \otimes f_q \right\rangle \\ &= \sum_{i,j,p,q} v_i^* w_j^* \tilde{v}_p \tilde{w}_q \langle e_i \otimes f_j, e_p \otimes f_q \rangle \\ &= \sum_{i,j,p,q} v_i^* w_j^* \tilde{v}_p \tilde{w}_q \delta_{ip} \delta_{jq} = \sum_i v_i^* \tilde{v}_i \sum_j w_j^* \tilde{w}_j \\ &= \langle v, \tilde{v} \rangle \langle w, \tilde{w} \rangle. \end{aligned} \quad (1.31)$$

The verification that the inner-product on $V \otimes W$ satisfies the remaining axioms is left as a good practice for you. Assume below that $X, Y \in V \otimes W$. For both exercises above simply write the most general vector, as $X = \sum_{ij} x_{ij} e_i \otimes f_j$ and proceed.

Exercise: Show that $\langle X, X \rangle \geq 0$, and $\langle X, X \rangle = 0$ if and only if $X = 0$.

Exercise: Show that $\langle X, Y \rangle = \langle Y, X \rangle^*$.

Many times it is convenient to use bra-ket notation for inner products in the tensor product. We write

$$\begin{aligned} |v \otimes w\rangle &= |v\rangle_1 \otimes |w\rangle_2 \\ \langle v \otimes w| &= {}_1\langle v|_1 \otimes {}_2\langle w|. \end{aligned} \quad (1.32)$$

Notice that both on bras and kets we write the state of particle one to the left of the state of particle two. We then write (1.28) as

$$\langle v \otimes w | \tilde{v} \otimes \tilde{w} \rangle = ({}_1\langle v| \otimes {}_2\langle w|) (|\tilde{v}\rangle_1 \otimes |\tilde{w}\rangle_2) = \langle v|\tilde{v}\rangle \langle w|\tilde{w}\rangle. \quad (1.33)$$

Back to our example with spin states, our four basis vectors $|+\rangle_1 \otimes |+\rangle_2$, $|+\rangle_1 \otimes |-\rangle_2$, $|-\rangle_1 \otimes |+\rangle_2$, and $|-\rangle_1 \otimes |-\rangle_2$ are orthonormal. We had the un-normalized state in (1.25) given by

$$|\Psi\rangle = \alpha \left(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2 \right). \quad (1.34)$$

The associated bra is then

$$\langle \Psi| = \alpha^* \left({}_1\langle +| \otimes {}_2\langle -| - {}_1\langle -| \otimes {}_2\langle +| \right). \quad (1.35)$$

We then have

$$\begin{aligned} \langle \Psi | \Psi \rangle &= \alpha \alpha^* \left({}_1\langle +| \otimes {}_2\langle -| - {}_1\langle -| \otimes {}_2\langle +| \right) \left(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2 \right) \\ &= \alpha \alpha^* \left({}_1\langle +| \otimes {}_2\langle -| |+\rangle_1 \otimes |-\rangle_2 + {}_1\langle -| \otimes {}_2\langle +| |-\rangle_1 \otimes |+\rangle_2 \right) \end{aligned} \quad (1.36)$$

since only terms where the spin states are the same for the first particle and for the second particle survive. We thus have, for normalization,

$$\langle \Psi | \Psi \rangle = |\alpha|^2 (1 + 1) = 2|\alpha|^2 = 1, \quad \rightarrow \quad \alpha = \frac{1}{\sqrt{2}}. \quad (1.37)$$

The normalized state with zero total angular momentum is then

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2 \right). \quad (1.38)$$

2 Entangled States

You have learned that $V \otimes W$ includes states $\Psi = \sum_i \alpha_i v_i \otimes w_i$ obtained by linear superposition of simpler states of the form $v_i \otimes w_i$. If handed such a Ψ , you might want to know whether you can write it as a single term $v_* \otimes w_*$ for some $v_* \in V$ and $w_* \in W$. If so, you are able to describe the state of the particles in Ψ independently: particle one is in state v_* and particle two in state w_* . We then say that in the state Ψ the particles are *not entangled*. If no such v_* and w_* exist, we say that in the state $\Psi \in V \otimes W$ the particles are entangled or equivalently, that Ψ is an entangled state of the two particles. Entanglement is a basis-independent property.

It is simplest to illustrate this using two-dimensional complex vector spaces V and W , like the ones we use for spin one-half. Let V have a basis e_1, e_2 and W have a basis f_1, f_2 . Then, the most general state you can write is the following:

$$\Psi_A = a_{11} e_1 \otimes f_1 + a_{12} e_1 \otimes f_2 + a_{21} e_2 \otimes f_1 + a_{22} e_2 \otimes f_2. \quad (2.39)$$

This state is encoded by a matrix A of coefficients

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \quad (2.40)$$

The state is *not entangled* if there exist constants a_1, a_2, b_1, b_2 such that

$$a_{11} e_1 \otimes f_1 + a_{12} e_1 \otimes f_2 + a_{21} e_2 \otimes f_1 + a_{22} e_2 \otimes f_2 = (a_1 e_1 + a_2 e_2) \otimes (b_1 f_1 + b_2 f_2). \quad (2.41)$$

Note that these four unknown constants are not uniquely determined: we can, for example, multiply a_1 and a_2 by some constant $c \neq 0$ and divide b_1 and b_2 by c , to obtain a different solution. Indeed $v \otimes w = (cv) \otimes (w/c)$ for any $c \neq 0$. Using the distributive laws for \otimes to expand the right-hand side of (2.41) and recalling that $e_i \otimes f_j$ are basis vectors in the tensor product, we see that the equality requires the following four relations:

$$\begin{aligned} a_{11} &= a_1 b_1 \\ a_{12} &= a_1 b_2 \\ a_{21} &= a_2 b_1 \\ a_{22} &= a_2 b_2 \end{aligned} \quad (2.42)$$

Combining these four expressions leaves us with a consistency condition:

$$a_{11}a_{22} - a_{12}a_{21} = a_1b_1a_2b_2 - a_1b_2a_2b_1 = 0 \rightarrow \det A = 0. \quad (2.43)$$

In other words, if Ψ_A is not entangled the determinant of the matrix A must be zero. We can in fact show that $\det A = 0$ implies that Ψ_A is not entangled. To do this we simply have to present a solution for the equations above under the condition $\det A = 0$.

Assume first that $a_{11} = 0$. Then $\det A = 0$ implies $a_{12}a_{21} = 0$. If $a_{12} = 0$ then

$$\Psi_A = a_{21}e_2 \otimes f_1 + a_{22}e_2 \otimes f_2 = e_2 \otimes (a_{21}f_1 + a_{22}f_2) \quad (2.44)$$

and the state is indeed not entangled. If $a_{21} = 0$ then

$$\Psi_A = a_{12}e_1 \otimes f_2 + a_{22}e_2 \otimes f_2 = (a_{12}e_1 + a_{22}e_2) \otimes f_2 \quad (2.45)$$

and again, the state is not entangled. Thus, we can solve all equations when $a_{11} = 0$. Now assuming $a_{11} \neq 0$ we can take

$$a_1 = \sqrt{a_{11}}, \quad b_1 = \sqrt{a_{11}}, \quad (2.46)$$

to solve the first equation in (2.42). The second and third equations allow us to solve for b_2 and a_2

$$b_2 = \frac{a_{12}}{\sqrt{a_{11}}}, \quad a_2 = \frac{a_{21}}{\sqrt{a_{11}}} \quad (2.47)$$

The fourth equation is then automatically satisfied as

$$a_2 b_2 = \frac{a_{12}a_{21}}{a_{11}} = \frac{a_{11}a_{22}}{a_{11}} = a_{22} \quad (2.48)$$

using the vanishing determinant condition. We have thus solved the system of equations and we can write

$$\Psi_A = \left(\sqrt{a_{11}}e_1 + \frac{a_{21}}{\sqrt{a_{11}}}e_2 \right) \otimes \left(\sqrt{a_{11}}f_1 + \frac{a_{12}}{\sqrt{a_{11}}}f_2 \right) \quad \text{if } \det A = 0. \quad (2.49)$$

We have thus proved that Ψ_A is entangled if and only if $\det A \neq 0$. For vector spaces of dimensions different than two the conditions for entanglement take a different form. Schrödinger called “entanglement” the essential feature of quantum mechanics.

Example: Consider our state of zero total spin angular momentum:

$$|\Phi\rangle_A \equiv \frac{1}{\sqrt{2}}(|+\rangle_1 \otimes |+\rangle_2 - |-\rangle_1 \otimes |-\rangle_2) \quad (2.50)$$

If we have the basis vectors $|e_1\rangle = |+\rangle_1, |e_2\rangle = |-\rangle_1$ and $|f_1\rangle = |+\rangle_2, |f_2\rangle = |-\rangle_2$ we see that the state is described by the matrix

$$A = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} \end{pmatrix} \quad (2.51)$$

Since the determinant of this matrix is not zero, the state is entangled.

3 Bell basis states

Bell states are a set of entangled basis vectors. Take $V_1 \otimes V_2$, with V_1 and V_2 both the two-dimensional complex vector space of spin-1/2 particles. For brevity of notation we will leave out the 1 and 2

subscripts on the states and the \otimes in between the states; it is always understood that in $V_1 \otimes V_2$ the state of V_1 appears to the left of the state of V_2 . Consider now the state

$$|\Phi_0\rangle \equiv \frac{1}{\sqrt{2}}(|+\rangle|+\rangle + |-\rangle|-\rangle). \quad (3.52)$$

This is clearly an entangled state: its associated matrix is diagonal with equal entries of $1/\sqrt{2}$ and thus non-zero determinant. Moreover this state is unit normalized

$$\langle \Phi_0 | \Phi_0 \rangle = 1. \quad (3.53)$$

We can use this state as the first of our basis vectors for $V_1 \otimes V_2$. Since this tensor product is four-dimensional we need three more entangled basis states. Here they are:

$$|\Phi_i\rangle \equiv (\mathbf{1} \otimes \sigma_i)|\Phi_0\rangle, \quad i = 1, 2, 3. \quad (3.54)$$

We will explicitly see below that these states are entangled, but this property is clear from the definition. If $|\Psi_i\rangle$ is not entangled, it would follow that that $\mathbf{1} \otimes \sigma_i|\Psi_i\rangle$ (i not summed) is not entangled either (do you see why?). But using $\sigma_i^2 = 1$, we see that this last state is in fact $|\Phi_0\rangle$, which is entangled. This contradiction shows that $|\Phi_i\rangle$ must be entangled. It is also manifest from the definition that the $|\Phi_i\rangle$ states are unit normalized.

Let us look at the form of $|\Phi_1\rangle$:

$$\begin{aligned} |\Phi_1\rangle &= (\mathbf{1} \otimes \sigma_1)\frac{1}{\sqrt{2}}(|+\rangle|+\rangle + |-\rangle|-\rangle) = \frac{1}{\sqrt{2}}(|+\rangle\sigma_1|+\rangle + |-\rangle\sigma_1|-\rangle) \\ &= \frac{1}{\sqrt{2}}(|+\rangle|-\rangle + |-\rangle|+\rangle). \end{aligned} \quad (3.55)$$

The state is clearly entangled. By analogous calculations we obtain the full list of Bell states

$$\begin{aligned} |\Phi_0\rangle &= \mathbf{1} \otimes \mathbf{1} |\Phi_0\rangle = \frac{1}{\sqrt{2}}(|+\rangle|+\rangle + |-\rangle|-\rangle) \\ |\Phi_1\rangle &= \mathbf{1} \otimes \sigma_1 |\Phi_0\rangle = \frac{1}{\sqrt{2}}(|+\rangle|-\rangle + |-\rangle|+\rangle) \\ |\Phi_2\rangle &= \mathbf{1} \otimes \sigma_2 |\Phi_0\rangle = \frac{i}{\sqrt{2}}(|+\rangle|-\rangle - |-\rangle|+\rangle) \\ |\Phi_3\rangle &= \mathbf{1} \otimes \sigma_3 |\Phi_0\rangle = \frac{1}{\sqrt{2}}(|+\rangle|+\rangle - |-\rangle|-\rangle). \end{aligned} \quad (3.56)$$

By inspection we can confirm that Φ_0 is orthogonal to the other three: $\langle \Phi_0 | \Phi_i \rangle = 0$. It is not much work either to see that the basis is in fact orthonormal. But a calculation is kind of fun:

$$\begin{aligned} \langle \Phi_i | \Phi_j \rangle &= \langle \Phi_0 | (\mathbf{1} \otimes \sigma_i)(\mathbf{1} \otimes \sigma_j) |\Phi_0\rangle \\ &= \langle \Phi_0 | \mathbf{1} \otimes \sigma_i \sigma_j |\Phi_0\rangle \\ &= \langle \Phi_0 | \mathbf{1} \otimes (\mathbf{1} \delta_{ij} + i \epsilon_{ijk} \sigma_k) |\Phi_0\rangle \\ &= \delta_{ij} \langle \Phi_0 | \mathbf{1} \otimes \mathbf{1} |\Phi_0\rangle + i \epsilon_{ijk} \langle \Phi_0 | \mathbf{1} \otimes \sigma_k |\Phi_0\rangle \\ &= \delta_{ij} \langle \Phi_0 | \Phi_0 \rangle + i \epsilon_{ijk} \langle \Phi_0 | \Phi_k \rangle = \delta_{ij}, \end{aligned} \quad (3.57)$$

as we wanted to show. Indeed, we have an orthonormal basis of entangled states.

We can solve for the old, non-entangled basis states in terms of the Bell states. We quickly find from (3.56)

$$\begin{aligned} |+\rangle|+\rangle &= \frac{1}{\sqrt{2}} (|\Phi_0\rangle + |\Phi_3\rangle) \\ |-\rangle|-\rangle &= \frac{1}{\sqrt{2}} (|\Phi_0\rangle - |\Phi_3\rangle) \\ |+\rangle|-\rangle &= \frac{1}{\sqrt{2}} (|\Phi_1\rangle - i|\Phi_2\rangle) \\ |-\rangle|+\rangle &= \frac{1}{\sqrt{2}} (|\Phi_1\rangle + i|\Phi_2\rangle). \end{aligned} \tag{3.58}$$

Introducing labels A and B for the two spaces in a tensor product $V_A \otimes V_B$ we rewrite the above equations as

$$\begin{aligned} |+\rangle_A|+\rangle_B &= \frac{1}{\sqrt{2}} (|\Phi_0\rangle_{AB} + |\Phi_3\rangle_{AB}) \\ |-\rangle_A|-\rangle_B &= \frac{1}{\sqrt{2}} (|\Phi_0\rangle_{AB} - |\Phi_3\rangle_{AB}) \\ |+\rangle_A|-\rangle_B &= \frac{1}{\sqrt{2}} (|\Phi_1\rangle_{AB} - i|\Phi_2\rangle_{AB}) \\ |-\rangle_A|+\rangle_B &= \frac{1}{\sqrt{2}} (|\Phi_1\rangle_{AB} + i|\Phi_2\rangle_{AB}), \end{aligned} \tag{3.59}$$

where $|\Phi_i\rangle_{AB}$ are the Bell states we defined above with tensor products in which the first state is in V_A and the second state is in V_B .

These basis states form the Bell basis. You could do an experiment to determine the probability of an arbitrary state being along any of the basis states in this orthonormal basis. You can use the experiment to detect which basis state the state is in. The state is, of course, a superposition of basis states, but during measurement will collapse into one of them with some probability. The Stern Gerlach device was an example of a device that allowed you to collapse a state into one basis state or another. This basis is more general, as it is not simply for two-state systems.

We conclude by presenting three facts.

1. Measuring in a basis. Given an orthonormal basis $|e_1\rangle, \dots, |e_n\rangle$ we can measure a state $|\Psi\rangle$ along this basis and obtain that the probability $P(i)$ to be in the state $|i\rangle$ is $|\langle e_i|\Psi\rangle|^2$. After measurement the state will be in one of the states $|e_i\rangle$. This is exactly how it worked for the Stern-Gerlach experiment which, oriented about **z** amount to a measurement in the basis $|+\rangle, |-\rangle$.

As another example, if we have a state with two particles A, B , we may choose the four Bell states as our orthonormal basis for the measurement. If so, after measurement the state will be in one of the Bell states $|\Phi_i\rangle_{AB}$, with probability given by the squared overlap $|\langle \Phi_i|_{AB}|\Psi\rangle|^2$.

2. Partial measurement. Suppose we have a general (entangled) state $\Psi \in V \otimes W$ of two particles. The observer Alice has access to both particles but decides to measure only the first particle along the basis $|e_1\rangle, \dots, |e_n\rangle$ of V . How is this analyzed? As a first step we use that basis to write the state Ψ in the form

$$\Psi = \sum_i |e_i\rangle \otimes |w_i\rangle, \quad (3.60)$$

for some calculable vectors $|w_i\rangle$. As a second step we normalize the states $|w_i\rangle$:

$$\Psi = \sum_i \sqrt{\langle w_i|w_i\rangle} |e_i\rangle \otimes \frac{|w_i\rangle}{\sqrt{\langle w_i|w_i\rangle}}, \quad (3.61)$$

We claim that Alice will find the first particle to be in the state $|i\rangle$ with probability $\langle w_i|w_i\rangle$. After the measurement, the state of the particles will be

$$|e_i\rangle \otimes \frac{|w_i\rangle}{\sqrt{\langle w_i|w_i\rangle}}, \quad \text{for some value of } i. \quad (3.62)$$

(A justification of this answer was given in recitations.) You probably have used this rule before. As an example, suppose we have the entangled state of total spin zero:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2) \quad (3.63)$$

If we measure the first particle along the $|+\rangle_1, |+\rangle_2$ basis we find

$$\begin{aligned} &\text{Probability that the first particle is in } |+\rangle = \frac{1}{2}. \quad \text{State after measurement: } |+\rangle_1 \otimes |-\rangle_2 \\ &\text{Probability that the first particle is in } |-\rangle = \frac{1}{2}. \quad \text{State after measurement: } |-\rangle_1 \otimes |+\rangle_2 \end{aligned} \quad (3.64)$$

It follows that after the measurement of the first particle, a measurement of the second particle will show that its spin is always opposite to the spin of the first particle.

As a more nontrivial example, consider now the state of three particles A, B, C which live in $V_A \otimes V_B \otimes V_C$, which contains states of the type $v \otimes w \otimes u$ with $v \in V_A, w \in V_B, u \in V_C$, and their linear combinations. To analyze what happens if Alice decides to do a Bell measurement of particles A, B , the state Ψ of the system must be written in the form

$$\Psi = |\Phi_0\rangle_{AB} \otimes |u_0\rangle_C + |\Phi_1\rangle_{AB} \otimes |u_1\rangle_C + |\Phi_2\rangle_{AB} \otimes |u_2\rangle_C + |\Phi_3\rangle_{AB} \otimes |u_3\rangle_C \quad (3.65)$$

After the measurement, the state of the particles A, B will be one of the Bell states $|\Phi_\mu\rangle_{AB}$ with $\mu = 0, 1, 2, 3$. We have

$$\begin{aligned} &\text{Probability that } (A, B) \text{ is in } |\Phi_\mu\rangle_{AB} = \langle u_\mu|u_\mu\rangle, \\ &\text{State after measurement is } |\Phi_\mu\rangle_{AB} \otimes \frac{|u_\mu\rangle_C}{\sqrt{\langle u_\mu|u_\mu\rangle}} \text{ for some } \mu \in \{0, 1, 2, 3\}. \end{aligned} \quad (3.66)$$

3. The action of the Pauli matrices on spin states can be realized as time evolution via some Hamiltonian. Note first that the Pauli matrices are unitary because they are Hermitian and square to the identity. Multiplying a state by σ_1 is thus acting with a unitary operator and unitary operators generate allowed time evolution. Thus, there is a Hamiltonian that applied to a system over some length of time will turn any spin state $|\Psi\rangle$ into $\sigma_i|\Psi\rangle$. In practice, this Hamiltonian would correspond to some device with a magnetic field of some determined magnitude and direction that acts for a few picoseconds and evolves spin states in time. We can check, for example, that any Pauli matrix can be written as the exponential of i times a Hermitian matrix (which would be proportional to the Hamiltonian):

$$e^{i\frac{\pi}{2}(-1+\sigma_i)} = e^{-i\frac{\pi}{2}e^{i\frac{\pi}{2}\sigma_i}} = (-i)(i\sigma_i) = \sigma_i \quad (3.67)$$

4 Quantum Teleportation

Classically, teleportation is impossible: there is no classical basis for dematerializing an object and recreating it somewhere else. In 1993, a group of scientists (Bennet, Brassard, Crépeau, Jozsa, Peres, and Wootters) discovered that teleportation *is* possible in quantum mechanics.

Imagine that Alice has a quantum state: the state of a 1/2 particle. The state is:

$$|\Psi\rangle_C = \alpha|+\rangle_C + \beta|-\rangle_C, \quad (4.68)$$

where $\alpha, \beta \in \mathbb{C}$ and the letter C denotes the state space V_C of this C particle to be teleported. Her goal is to teleport this state - called a “quantum bit,” or *qubit* - to Bob, who is far away.

The quantum “no-cloning” principle prevents Alice from simply creating a copy of the state and sending it to Bob. In other words, it is impossible to create a copy of a quantum mechanical state. Measuring the state and telling Bob about the result is no option either: if Alice measures the state with some Stern-Gerlach apparatus, the spin will just point up or point down. What has she learned? Almost nothing. Only with many copies of the state she would be able to learn about the values of α and β . Having just one particle she is unable to measure α and β and send those values to Bob. Of course, it may be that for some reason Alice knows the values of α and β . In some cases she could transmit that information to Bob to recreate the state. But it could also be that α , for example is some transcendental number 0.178573675623..... with no discernible rhyme or reason, and she would need an infinite amount of information to send to Bob this value.

Here is a diagram of how Alice can will teleport the information:

The key tool Alice and Bob use is an entangled states of two particles A and B , in which Alice has access to particle A and Bob has access to particle B . One pair (A, B) of entangled particles will allow Alice to teleport the state C of one particle. To teleport a full person from one place to another, we would have to have an enormous reservoir of entangled pairs, one pair needed to teleport each quantum state of particles in the body of that person. This clearly remains science-fiction.

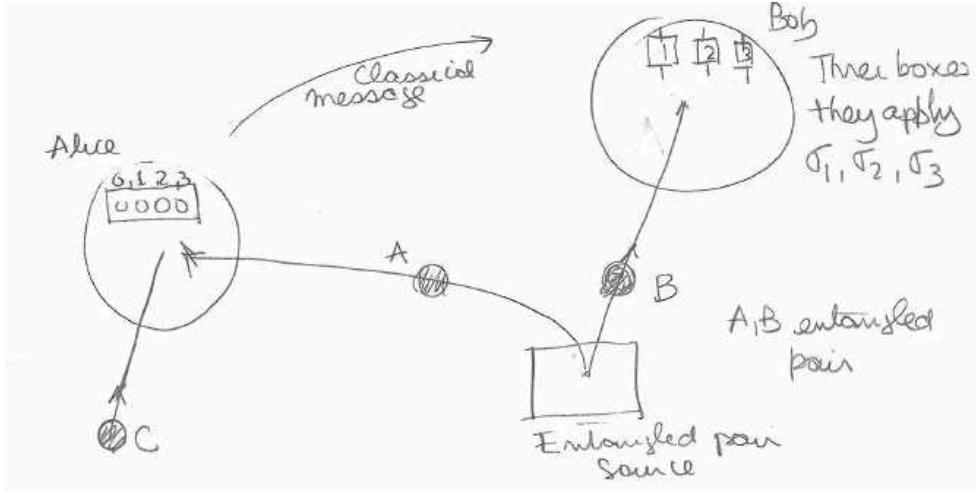


Figure 1: Alice has access to spin state C , to be teleported and to spin state A which is entangled with spin state B . Bob has spin state B . Alice will perform a Bell measurement on states A and C . After she measures, Bob's state will turn into the desired state, up to a simple unitary transformation.

Alice has a console with four lights, labeled with $\mu = 0, 1, 2, 3$. She will do a Bell measurement involving particle A (of the shared entangled pair) and particle C (the one to be teleported). As she does so, one of her four lights will blink: If it is the μ -th light it is because she ended with the Bell state $|\Phi_\mu\rangle_{AC}$. Bob, who is in possession of the particle B , has a console with four boxes which generate unitary transformations (via some Hamiltonians applied for a fixed time). The first box, labeled $\mu = 0$ does nothing to the state. The i -th box (with $i = 1, 2, 3$) applies the operator σ_i . Alice communicates to Bob that the μ -th light blinked. Then Bob submits particle B to the μ -th box and out comes, we claim, the teleported state $|\Psi\rangle_B$ as the state of particle B .

Let us prove this mathematically. Let the entangled shared pair, with A at Alice and B at Bob, be the first Bell basis state:

$$|\Phi_0\rangle_{AB} = \frac{1}{\sqrt{2}}(|+\rangle_A|+\rangle_B + |-\rangle_A|-\rangle_B). \quad (4.69)$$

The total state of our three particles, A, B, C is therefore:

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= |\Phi_0\rangle_{AB} \otimes (\alpha|+\rangle_C + \beta|-\rangle_C) \\ &= \frac{1}{\sqrt{2}}\left(|+\rangle_A|+\rangle_B + |-\rangle_A|-\rangle_B\right) \otimes (\alpha|+\rangle_C + \beta|-\rangle_C). \end{aligned} \quad (4.70)$$

Expanding out and reordering the states to have A followed by C and then by B we have

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= \frac{1}{\sqrt{2}}\left(\underbrace{\alpha|+\rangle_A|+\rangle_C}_{\text{underbrace}}|+\rangle_B + \underbrace{\beta|+\rangle_A|-\rangle_C}_{\text{underbrace}}|+\rangle_B \right. \\ &\quad \left. + \underbrace{\alpha|-\rangle_A|+\rangle_C}_{\text{underbrace}}|-\rangle_B + \underbrace{\beta|-\rangle_A|-\rangle_C}_{\text{underbrace}}|-\rangle_B\right). \end{aligned} \quad (4.71)$$

Note that as long as we label the states, the order in which we write them does not matter! We now write these basis states with braces in the Bell basis using (3.59). We find

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= \frac{1}{2}(|\Phi_0\rangle_{AC} + |\Phi_3\rangle_{AC}) \alpha|+\rangle_B + \frac{1}{2}(|\Phi_1\rangle_{AC} - i|\Phi_2\rangle_{AC}) \beta|+\rangle_B \\ &\quad + \frac{1}{2}(|\Phi_1\rangle_{AC} + i|\Phi_2\rangle_{AC}) \alpha|-\rangle_B + \frac{1}{2}(|\Phi_0\rangle_{AC} - |\Phi_3\rangle_{AC}) \beta|-\rangle_B. \end{aligned} \quad (4.72)$$

Collecting the Bell states we find

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= \frac{1}{2}|\Phi_0\rangle_{AC} (\alpha|+\rangle_B + \beta|-\rangle_B) + \frac{1}{2}|\Phi_1\rangle_{AC} (\alpha|-\rangle_B + \beta|+\rangle_B) \\ &\quad + \frac{1}{2}|\Phi_2\rangle_{AC} (i\alpha|-\rangle_B - i\beta|+\rangle_B) + \frac{1}{2}|\Phi_3\rangle_{AC} (\alpha|+\rangle_B - \beta|-\rangle_B). \end{aligned} \quad (4.73)$$

We can then see that in fact we got

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= \frac{1}{2}|\Phi_0\rangle_{AC} \otimes |\Psi\rangle_B + \frac{1}{2}|\Phi_1\rangle_{AC} \otimes \sigma_1|\Psi\rangle_B \\ &\quad + \frac{1}{2}|\Phi_2\rangle_{AC} \otimes \sigma_2|\Psi\rangle_B + \frac{1}{2}|\Phi_3\rangle_{AC} \otimes \sigma_3|\Psi\rangle_B. \end{aligned} \quad (4.74)$$

The above right-hand side allows us to understand what happens when Alice measures the state of (A, C) in the Bell basis. If she measures:

- $|\Phi_0\rangle_{AC}$, then the B state becomes $|\Psi\rangle_B$,
- $|\Phi_1\rangle_{AC}$, then the B state becomes $\sigma_1|\Psi\rangle_B$,
- $|\Phi_2\rangle_{AC}$, then the B state becomes $\sigma_2|\Psi\rangle_B$,
- $|\Phi_3\rangle_{AC}$, then the B state becomes $\sigma_3|\Psi\rangle_B$.

If Alice got $|\Phi_0\rangle_{AC}$ then Bob is in the possession of the teleported state and has to do nothing. If Alice gets $|\Phi_i\rangle_{AC}$, Bob's particle is goes into the state $\sigma_i|\Psi\rangle_B$. Bob applies the i -th box, which multiplies his state by σ_i giving him the desired state $|\Psi\rangle_B$. The teleporting is thus complete!

Note that Alice is left with one of the Bell states $|\Phi_\mu\rangle_{AC}$ which has no information whatsoever about the constants α and β that defined the state to be teleported. Thus the process did not create a copy of the state. The original state is destroyed in the process of teleportation.

It is noteworthy that all the mathematical work above led to the key result (4.74), which is neatly summarized as the following identity valid for arbitrary states $|\Psi\rangle$:

$$|\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C = \frac{1}{2} \sum_{i=0}^3 |\Phi_i\rangle_{AC} \otimes \sigma_i|\Psi\rangle_B. \quad (4.75)$$

This is an identity for a state of three particles. It expresses the tensor product of an entangled state of the first two particles, times a third, as a sum of products that involve entangled states of the first and third particle times a state of the second particle.

5 EPR and Bell Inequalities

In this section we begin by studying some properties of the singlet state of two particles of spin-1/2. We then turn to the claims of Einstein, Podolsky, and Rosen (EPR) concerning quantum mechanics. Finally, we discuss the so-called Bell inequalities that would follow if EPR were right. Of course, quantum mechanics violates these inequalities, and experiment shows that the inequalities are indeed violated. EPR were wrong.

We have been talking about the singlet state of two spin-1/2 particles. This state emerges, for example, in particle decays. The neutral η_0 meson (of rest mass 547 MeV) sometimes decays into two oppositely charged muons

$$\eta_0 \rightarrow \mu^+ + \mu^- . \quad (5.1)$$

The meson is a spinless particle and being at rest has zero orbital angular momentum. As a result it has zero total angular momentum. As it decays, the final state of the two muons must have zero total angular momentum as well. If the state of the two muons has zero orbital angular momentum, it must also have zero total spin angular momentum. The two muons flying away from each other with zero orbital angular momentum are in a singlet state. This state takes the form

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|+\rangle_1|-\rangle_2 - |-\rangle_1|+\rangle_2) . \quad (5.2)$$

This singlet state is rotational invariant and therefore it is actually the same for whatever choice of direction \mathbf{n} to define a basis of spin states:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\mathbf{n};+\rangle_1|\mathbf{n};-\rangle_2 - |\mathbf{n};-\rangle_1|\mathbf{n};+\rangle_2) . \quad (5.3)$$

We now ask: In this singlet, what is the probability $P(\mathbf{a}, \mathbf{b})$ that the first particle is in the state $|\mathbf{a}; +\rangle$ and the second particle is in the state $|\mathbf{b}; +\rangle$, with \mathbf{a} and \mathbf{b} two arbitrarily chosen unit vectors? To help ourselves, we write the singlet state using the first vector

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a};+\rangle_1|\mathbf{a};-\rangle_2 - |\mathbf{a};-\rangle_1|\mathbf{a};+\rangle_2) . \quad (5.4)$$

By definition, the probability we want is

$$P(\mathbf{a}, \mathbf{b}) = \left| {}_1\langle \mathbf{a}; +| {}_2\langle \mathbf{b}; +|\Psi\rangle \right|^2 \quad (5.5)$$

Only the first term in (5.4) contributes and we get

$$P(\mathbf{a}, \mathbf{b}) = \frac{1}{2}|\langle \mathbf{b}; +|\mathbf{a}; -\rangle|^2 \quad (5.6)$$

We recall that the overlap-squared between two spin states is given by the cosine-squared of half the angle in between them. Using figure 2 we see that the angle between \mathbf{b} and $-\mathbf{a}$ is $\pi - \theta_{ab}$, where θ_{ab} is the angle between \mathbf{b} and \mathbf{a} . Therefore

$$P(\mathbf{a}, \mathbf{b}) = \frac{1}{2}\cos^2\left(\frac{1}{2}(\pi - \theta_{ab})\right) \quad (5.7)$$

Our final result is therefore

$$P(\mathbf{a}, \mathbf{b}) = \frac{1}{2} \sin^2\left(\frac{1}{2}\theta_{ab}\right). \quad (5.8)$$

As a simple consistency check, if $\mathbf{b} = -\mathbf{a}$ then $\theta_{ab} = \pi$ and $P(\mathbf{a}, -\mathbf{a}) = 1/2$ which is what we expect.

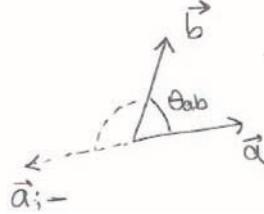


Figure 2: Directions associated with the vectors \mathbf{a} and \mathbf{b} .

If we measure about orthogonal vectors, like the unit vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ we get

$$P(\hat{\mathbf{z}}, \hat{\mathbf{x}}) = \frac{1}{2} \sin^2 45^\circ = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \quad (5.9)$$

The key statement of Einstein, Podolsky and Rosen (EPR) is the claim for **local realism**. This is posed as two properties of measurement:

1. The result of a measurement corresponds to some element of reality. If a measurement of an observable gives a value, that value was a property of the state.
2. The result of a measurement at one point cannot depend on whatever action takes place at a far away point at the same time.

Both properties seem quite plausible at first thought. The first, we are by now accustomed, is violated in Quantum Mechanics, where measurement involves collapse of the wavefunction, so that the result was not pre-ordained and does not correspond to a unequivocal property of the system. The violation of the second is perhaps equally disturbing, given our intuition that simultaneous spatially separated events can't affect each other. There is something non-local about quantum mechanics.

According to EPR the so called entangled pairs are just pairs of particles that have definite spins. They point out that the results of quantum mechanical measurements are reproduced if our large ensemble of pairs has the following distribution of states:

- In 50% of pairs, particle 1 has spin along $\hat{\mathbf{z}}$ and particle 2 has spin along $-\hat{\mathbf{z}}$,
- In 50% of pairs, particle 1 has spin along $-\hat{\mathbf{z}}$ and particle 2 has spin along $\hat{\mathbf{z}}$.

This would explain the perfect correlations and is consistent, for example, with $P(\hat{\mathbf{z}}, -\hat{\mathbf{z}}) = 1/2$, which we obtained quantum mechanically.

The challenge for EPR is to keep reproducing the results of more complicated measurements. Suppose each of the two observers can measure spin along two possible axes: the x and z axes. They measure once, in any of these two directions. EPR then state that in any pair each particle has a definite state of spin in these two directions. For example, a particle of type $(\hat{\mathbf{z}}, -\hat{\mathbf{x}})$ is one that if measured along z gives a plus $\hbar/2$ and if measured along x gives $-\hbar/2$. We do not do simultaneous measurements or subsequent measurements on each particle. EPR then claim that the observed quantum mechanical results are matched if our ensemble of pairs have the following properties

- 25% of pairs have particle 1 in $(\hat{\mathbf{z}}, \hat{\mathbf{x}})$ and particle 2 in $(-\hat{\mathbf{z}}, -\hat{\mathbf{x}})$
- 25% of pairs have particle 1 in $(\hat{\mathbf{z}}, -\hat{\mathbf{x}})$ and particle 2 in $(-\hat{\mathbf{z}}, \hat{\mathbf{x}})$
- 25% of pairs have particle 1 in $(-\hat{\mathbf{z}}, \hat{\mathbf{x}})$ and particle 2 in $(\hat{\mathbf{z}}, -\hat{\mathbf{x}})$
- 25% of pairs have particle 1 in $(-\hat{\mathbf{z}}, -\hat{\mathbf{x}})$ and particle 2 in $(\hat{\mathbf{z}}, \hat{\mathbf{x}})$

First note the complete correlations: particles one and two have opposite spins in each possible direction. This is, of course, needed to match the quantum mechanical singlets. We can ask what is $P(\hat{\mathbf{z}}, -\hat{\mathbf{z}})$, the probability that particle one is along $\hat{\mathbf{z}}$ and particle two along $-\hat{\mathbf{z}}$. The first two cases above apply, and thus this probability is $1/2$, consistent with quantum mechanics. We can also ask for $P(\hat{\mathbf{z}}, \hat{\mathbf{x}})$. This time only the second case applies giving us a probability of $1/4$ as we obtained earlier in (5.9). The quantum mechanical answers indeed arise.

The insight of Bell was that with Stern-Gerlach apparatuses that could measure in three directions one gets in trouble. Suppose each observer can measure along any one of the three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$. Again, each particle is just measured once. Let us assume that we have a large number N of pairs that, following EPR, contain particles with well-defined spins on these three directions. A particle of type $(\mathbf{a}, -\mathbf{b}, \mathbf{c})$, for example, if measured along \mathbf{a} would give $\hbar/2$, if measured along \mathbf{b} would give $-\hbar/2$ and if measured along \mathbf{c} would give $\hbar/2$. The following distribution is given:

Populations	Particle 1	Particle 2
N_1	$(\mathbf{a}, \mathbf{b}, \mathbf{c})$	$(-\mathbf{a}, -\mathbf{b}, -\mathbf{c})$
N_2	$(\mathbf{a}, \mathbf{b}, -\mathbf{c})$	$(-\mathbf{a}, -\mathbf{b}, \mathbf{c})$
N_3	$(\mathbf{a}, -\mathbf{b}, \mathbf{c})$	$(-\mathbf{a}, \mathbf{b}, -\mathbf{c})$
N_4	$(\mathbf{a}, -\mathbf{b}, -\mathbf{c})$	$(-\mathbf{a}, \mathbf{b}, \mathbf{c})$
N_5	$(-\mathbf{a}, \mathbf{b}, \mathbf{c})$	$(\mathbf{a}, -\mathbf{b}, -\mathbf{c})$
N_6	$(-\mathbf{a}, \mathbf{b}, -\mathbf{c})$	$(\mathbf{a}, -\mathbf{b}, \mathbf{c})$
N_7	$(-\mathbf{a}, -\mathbf{b}, \mathbf{c})$	$(\mathbf{a}, \mathbf{b}, -\mathbf{c})$
N_8	$(-\mathbf{a}, -\mathbf{b}, -\mathbf{c})$	$(\mathbf{a}, \mathbf{b}, \mathbf{c})$

As required, all spins are correlated in particles one and two. We also have $N = \sum_{i=1}^8 N_i$. We now record the following probabilities that follow by inspection of the table:

$$P(\mathbf{a}, \mathbf{b}) = \frac{N_3 + N_4}{N}, \quad P(\mathbf{a}, \mathbf{c}) = \frac{N_2 + N_4}{N}, \quad P(\mathbf{c}, \mathbf{b}) = \frac{N_3 + N_7}{N}. \quad (5.10)$$

Consider now the trivially correct inequality:

$$N_3 + N_4 \leq N_3 + N_7 + N_2 + N_4, \quad (5.11)$$

that on account of (5.10) implies the **Bell inequality**

$$\boxed{P(\mathbf{a}, \mathbf{b}) \leq P(\mathbf{a}, \mathbf{c}) + P(\mathbf{c}, \mathbf{b}).} \quad (5.12)$$

If true quantum mechanically, given (5.8) we would have

$$\frac{1}{2} \sin^2 \frac{1}{2}\theta_{ab} \leq \frac{1}{2} \sin^2 \frac{1}{2}\theta_{ac} + \frac{1}{2} \sin^2 \frac{1}{2}\theta_{cb}. \quad (5.13)$$

But this is violated for many choices of angles. Take, for example, the planar configuration in Fig. 3:

$$\theta_{ab} = 2\theta, \quad \theta_{ac} = \theta_{cb} = \theta. \quad (5.14)$$

For this situation, the inequality becomes

$$\frac{1}{2} \sin^2 \theta \leq \sin^2 \frac{1}{2}\theta. \quad (5.15)$$

This fails for sufficiently small θ : $\frac{1}{2}\theta^2 \leq \frac{\theta^2}{4}$ is just plain wrong. In fact, the inequality goes wrong for any $\theta < \frac{\pi}{2}$. Experimental results have confirmed that Bell inequalities are violated and thus the original claim of local realism by EPR is wrong.

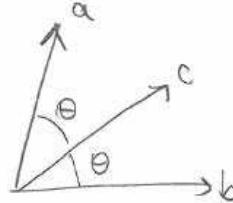


Figure 3: Special configuration for vectors \mathbf{a} , \mathbf{b} and \mathbf{c} .

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

UNCERTAINTY PRINCIPLE AND COMPATIBLE OBSERVABLES

B. Zwiebach

October 21, 2013

Contents

1 Uncertainty defined	1
2 The Uncertainty Principle	3
3 The Energy-Time uncertainty	6
4 Lower bounds for ground state energies	9
5 Diagonalization of Operators	11
6 The Spectral Theorem	12
7 Simultaneous Diagonalization of Hermitian Operators	16
8 Complete Set of Commuting Observables	18

1 Uncertainty defined

As we know, observables are associated to Hermitian operators. Given one such operator A we can use it to measure some property of the physical system, as represented by a state Ψ . If the state is in an eigenstate of the operator A , we have no uncertainty in the value of the observable, which coincides with the eigenvalue corresponding to the eigenstate. We only have uncertainty in the value of the observable if the physical state is not an eigenstate of A , but rather a superposition of various eigenstates with different eigenvalues.

We want to define the **uncertainty** $\Delta A(\Psi)$ of the Hermitian operator A on the state Ψ . This uncertainty should vanish if and only if the state is an eigenstate of A . The uncertainty, moreover, should be a real number. In order to define such uncertainty we first recall that the expectation value of A on the state Ψ , assumed to be normalized, is given by

$$\langle A \rangle = \langle \Psi | A | \Psi \rangle = \langle \Psi, A \Psi \rangle. \quad (1.1)$$

The expectation $\langle A \rangle$ is guaranteed to be real since A is Hermitian. We then define the uncertainty as the norm of the vector obtained by acting with $(A - \langle A \rangle I)$ on the physical state (I is the identity operator):

$$\boxed{\Delta A(\Psi) \equiv \left| (A - \langle A \rangle I) \Psi \right|.} \quad (1.2)$$

The uncertainty, so defined is manifestly non-negative. If the uncertainty is zero, the vector inside the norm is zero and therefore:

$$\Delta A(\Psi) = 0 \rightarrow (A - \langle A \rangle I)\Psi = 0 \rightarrow A\Psi = \langle A \rangle \Psi, \quad (1.3)$$

and the last equation confirms that the state is indeed an eigenstate of A (note that $\langle A \rangle$ is a number). You should also note that $\langle A \rangle$ is indeed the eigenvalue, since taking the eigenvalue equation $A\Psi = \lambda\Psi$ and forming the inner product with another Ψ we get

$$\langle \Psi, A\Psi \rangle = \lambda \langle \Psi, \Psi \rangle = \lambda \rightarrow \lambda = \langle A \rangle. \quad (1.4)$$

Alternatively, if the state Ψ is an eigenstate, we now know that the eigenvalue is $\langle A \rangle$ and therefore the state $(A - \langle A \rangle I)\Psi$ vanishes and its norm is zero. We have therefore shown that

The uncertainty $\Delta A(\Psi)$ vanishes if and only if Ψ is an eigenstate of A . (1.5)

To compute the uncertainty one usually squares the expression in (1.2) so that

$$(\Delta A(\Psi))^2 = \langle (A - \langle A \rangle I)\Psi, (A - \langle A \rangle I)\Psi \rangle \quad (1.6)$$

Since the operator A is assumed to be Hermitian and consequently $\langle A \rangle$ is real, we have $(A - \langle A \rangle I)^\dagger = A - \langle A \rangle I$, and therefore we can move the operator on the first entry onto the second one to find

$$(\Delta A(\Psi))^2 = \langle \Psi, (A - \langle A \rangle I)^2 \Psi \rangle. \quad (1.7)$$

While this is a reasonable form, we can simplify it further by expansion

$$(\Delta A(\Psi))^2 = \langle \Psi, (A^2 - 2\langle A \rangle A + \langle A \rangle^2 I)\Psi \rangle. \quad (1.8)$$

The last two term combine and we find

$$(\Delta A(\Psi))^2 = \langle A^2 \rangle - \langle A \rangle^2. \quad (1.9)$$

Since the left-hand side is greater than or equal to zero, this incidentally shows that the expectation value of A^2 is larger than the expectation value of A , squared:

$$\langle A^2 \rangle \geq \langle A \rangle^2. \quad (1.10)$$

An interesting geometrical interpretation of the uncertainty goes as follows. Consider the one-dimensional vector subspace U_Ψ generated by Ψ . Take the state $A\Psi$ and project it to the subspace U_Ψ . The projection, we claim is $\langle A \rangle \Psi$ and the part of $A\Psi$ in the orthogonal subspace U_Ψ^\perp is a vector of norm equal to the uncertainty ΔA . Indeed the orthogonal projector P_{U_Ψ} is

$$P_{U_\Psi} = |\Psi\rangle\langle\Psi|, \quad (1.11)$$

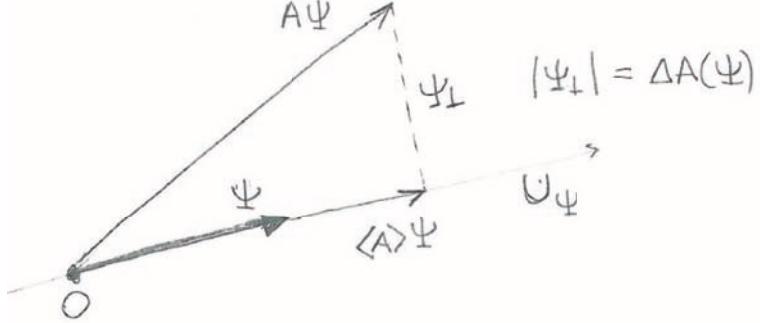


Figure 1: A state Ψ and the one-dimensional subspace U_Ψ generated by it. The projection of $A\Psi$ to U_Ψ is $\langle A \rangle \Psi$. The orthogonal complement Ψ_\perp is a vector whose norm is the uncertainty $\Delta A(\Psi)$.

so that

$$P_{U_\Psi} A|\Psi\rangle = |\Psi\rangle\langle\Psi|A|\Psi\rangle = |\Psi\rangle\langle A|. \quad (1.12)$$

Moreover, the vector $A|\Psi\rangle$ minus its projection must be a vector $|\Psi_\perp\rangle$ orthogonal to $|\Psi\rangle$

$$A|\Psi\rangle - \langle A|\Psi\rangle = |\Psi_\perp\rangle, \quad (1.13)$$

as is easily confirmed by taking the overlap with the bra Ψ . Since the norm of the above left-hand side is the uncertainty, we confirm that $\Delta A = |\Psi_\perp|$, as claimed. These results are illustrated in Figure 1.

2 The Uncertainty Principle

The uncertainty principle is an inequality that is satisfied by the product of the uncertainties of two Hermitian operators that fail to commute. Since the uncertainty of an operator on any given physical state is a number greater than or equal to zero, the product of uncertainties is also a real number greater than or equal to zero. The uncertainty inequality often gives us a lower bound for this product.

When the two operators in question commute, the uncertainty inequality gives no information.

Let us state the uncertainty inequality. Consider two Hermitian operators A and B and a physical state Ψ of the quantum system. Let ΔA and ΔB denote the uncertainties of A and B , respectively, in the state Ψ . Then we have

$$(\Delta A)^2(\Delta B)^2 \geq \left(\langle \Psi | \frac{1}{2i}[A, B] | \Psi \rangle \right)^2. \quad (2.14)$$

The left hand side is a real, non-negative number. For this to be consistent inequality, the right-hand side must also be a real number that is not negative. Since the right-hand side appears squared, the object inside the parenthesis must be real. This can only happen for all Ψ if the operator

$$\frac{1}{2i}[A, B] \quad (2.15)$$

is Hermitian. For this first note that the commutator of two Hermitian operators is *anti*-Hermitian:

$$[A, B]^\dagger = (AB)^\dagger - (BA)^\dagger = B^\dagger A^\dagger - A^\dagger B^\dagger - BA = -[A, B] \quad (2.16)$$

The presence of the i then makes the operator in (2.15) Hermitian. Note that the uncertainty inequality can also be written as

$$\Delta A \Delta B \geq \left| \langle \Psi | \frac{1}{2i} [A, B] | \Psi \rangle \right|. \quad (2.17)$$

where the bars on the right-hand side denote absolute value.

Before we prove the theorem, let's do the canonical example! Substituting \hat{x} for A and \hat{p} for B results in the position-momentum uncertainty relation you have certainly worked with:

$$(\Delta x)^2 (\Delta p)^2 \geq \left(\langle \Psi | \frac{1}{2i} [\hat{x}, \hat{p}] | \Psi \rangle \right)^2. \quad (2.18)$$

Since $[\hat{x}, \hat{p}]/(2i) = \hbar/2$ we get

$$(\Delta x)^2 (\Delta p)^2 \geq \frac{\hbar^2}{4} \rightarrow \Delta x \Delta p \geq \frac{\hbar}{2}. \quad (2.19)$$

We are interested in the proof of the uncertainty inequality for it gives the information that is needed to find the conditions that lead to saturation.

Proof. We define the following two states:

$$\begin{aligned} |f\rangle &\equiv (A - \langle A \rangle I)|\Psi\rangle \\ |g\rangle &\equiv (B - \langle B \rangle I)|\Psi\rangle. \end{aligned} \quad (2.20)$$

Note that by the definition (1.2) of uncertainty,

$$\begin{aligned} \langle f|f\rangle &= (\Delta A)^2, \\ \langle g|g\rangle &= (\Delta B)^2. \end{aligned} \quad (2.21)$$

The Schwarz inequality immediately furnishes us an inequality involving precisely the uncertainties

$$\langle f|f\rangle \langle g|g\rangle \geq |\langle f|g\rangle|^2, \quad (2.22)$$

and therefore we have

$$(\Delta A)^2 (\Delta B)^2 \geq |\langle f|g\rangle|^2 = (\text{Re}\langle f|g\rangle)^2 + (\text{Im}\langle f|g\rangle)^2. \quad (2.23)$$

Writing $\check{A} = (A - \langle A \rangle I)$ and $\check{B} = (B - \langle B \rangle I)$, we now begin to compute the right-hand side:

$$\langle f|g\rangle = \langle \Psi | \check{A} \check{B} | \Psi \rangle = \langle \Psi | (A - \langle A \rangle I)(B - \langle B \rangle I) | \Psi \rangle = \langle \Psi | AB | \Psi \rangle - \langle A \rangle \langle B \rangle, \quad (2.24)$$

and since $|f\rangle$ and $|g\rangle$ go into each other as we exchange A and B ,

$$\langle g|f\rangle = \langle \Psi | \check{A} \check{B} | \Psi \rangle = \langle \Psi | BA | \Psi \rangle - \langle B \rangle \langle A \rangle. \quad (2.25)$$

From the two equations above we find a nice expression for the imaginary part of $\langle f|g \rangle$:

$$\text{Im}\langle f|g \rangle = \frac{1}{2i}(\langle f|g \rangle - \langle g|f \rangle) = \frac{1}{2i}\langle \Psi|[A, B]|\Psi \rangle. \quad (2.26)$$

For the real part the expression is not that simple, so it is best to leave it as the anticommutator of the checked operators:

$$\text{Re}\langle f|g \rangle = \frac{1}{2}(\langle f|g \rangle + \langle g|f \rangle) = \frac{1}{2}\langle \Psi|\{\hat{A}, \hat{B}\}|\Psi \rangle \quad (2.27)$$

Back in (2.23) we get

$$(\Delta A)^2(\Delta B)^2 \geq \left(\langle \Psi|\frac{1}{2i}[A, B]|\Psi \rangle\right)^2 + \left(\langle \Psi|\frac{1}{2}\{\hat{A}, \hat{B}\}|\Psi \rangle\right)^2. \quad (2.28)$$

This can be viewed as the most complete form of the uncertainty inequality. It turns out, however, that the second term on the right hand side is seldom simple enough to be of use, and many times it can be made equal to zero for certain states. At any rate, the term is positive or zero so it can be dropped while preserving the inequality. This is often done, thus giving the celebrated form (2.14) that we have now established.

Now that we have a proven the uncertainty inequality, we can ask: What are the conditions for this inequality to be saturated? If the goal is to minimize uncertainties, under what conditions can we achieve the minimum possible product of uncertainties? As the proof shows, saturation is achieved under two conditions:

1. The Schwarz inequality is saturated. For this we need $|g\rangle = \beta|f\rangle$ where $\beta \in \mathbb{C}$.
2. $\text{Re}(\langle f|g \rangle) = 0$, so that the last term in (2.28) vanishes. This means that $\langle f|g \rangle + \langle g|f \rangle = 0$.

Using $|g\rangle = \beta|f\rangle$ in Condition 2, we get

$$\langle f|g \rangle + \langle g|f \rangle = \beta\langle f|f \rangle + \beta^*\langle f|f \rangle = (\beta + \beta^*)\langle f|f \rangle = 0, \quad (2.29)$$

which requires $\beta + \beta^* = 0$ or that the real part of β vanish. It follows that β must be purely imaginary. So, $\beta = i\lambda$, with λ real, and therefore the uncertainty inequality will be saturated if and only if

$$|g\rangle = i\lambda|f\rangle, \quad \lambda \in \mathbb{R}. \quad (2.30)$$

More explicitly this requires

$$\text{Saturation Condition: } (B - \langle B \rangle I)|\Psi \rangle = i\lambda(A - \langle A \rangle I)|\Psi \rangle. \quad (2.31)$$

This must be viewed as a condition for Ψ , given any two operators A and B . Moreover, note that $\langle A \rangle$ and $\langle B \rangle$ are Ψ dependent. What is λ , physically? Well, the norm of λ is actually fixed by the equation. Taking the norm of both sides we get

$$\Delta B = |\lambda|\Delta A \rightarrow |\lambda| = \frac{\Delta B}{\Delta A}. \quad (2.32)$$

The classic illustration of this saturation condition is worked out for the x, p uncertainty inequality $\Delta x \Delta p \geq \hbar/2$. You will find that gaussian wavefunctions satisfy the saturation condition.

3 The Energy-Time uncertainty

A more subtle form of the uncertainty relation deals with energy and time. The inequality is sometimes stated vaguely in the form $\Delta E \Delta t \gtrsim \hbar$. In here there is no problem in defining ΔE precisely, after all we have the Hamiltonian operator, and its uncertainty ΔH is a perfect candidate for the ‘energy uncertainty’. The problem is time. Time is not an operator in quantum mechanics, it is a parameter, a real number used to describe the way systems change. Unless we define Δt in a precise way we cannot hope for a well-defined uncertainty relation.

We can try a rough, heuristic definition, in order to illustrate the spirit of the inequality. Consider a photon that is detected at some point in space, as a passing oscillatory wave of exact duration T . Even without quantum mechanical considerations we can ask the observer what was the angular frequency ω of the pulse. In order to answer our question the observer will attempt to count the number N of complete oscillations of the waveform that went through. Of course, this number N is given by T divided by the period $2\pi/\omega$ of the wave:

$$N = \frac{\omega T}{2\pi}. \quad (3.33)$$

The observer, however, will typically fail to count full waves, because as the pulse gets started from zero and later on dies off completely, the waveform will cease to follow the sinusoidal pattern. Thus we expect an uncertainty $\Delta N \gtrsim 1$. Given the above relation, this implies an uncertainty $\Delta\omega$ in the value of the angular frequency

$$\Delta\omega T \gtrsim 2\pi. \quad (3.34)$$

This is all still classical, the above identity is something electrical engineers are well aware of. It represents a limit on the ability to ascertain accurately the frequency of a wave that is observed for a limited amount of time. This becomes quantum mechanical if we speak of a single photon, whose energy is $E = \hbar\omega$. Then $\Delta E = \hbar\Delta\omega$, so that multiplying the above inequality by \hbar we get

$$\Delta E T \gtrsim \hbar. \quad (3.35)$$

In this uncertainty inequality T is the duration of the pulse. It is a reasonable relation but the presence of \gtrsim betrays its lack of full precision.

We can find a precise energy/ Q -ness uncertainty inequality by applying the general uncertainty inequality to the Hamiltonian H and another Hermitian operator Q , as did the distinguished Russian physicists L. Mandelstam and Tamm shortly after the formulation of the uncertainty principle. We would then have

$$\Delta H \Delta Q \geq \left| \langle \Psi | \frac{1}{2i} [H, Q] | \Psi \rangle \right|. \quad (3.36)$$

This starting point is interesting because the commutator $[H, Q]$ encodes something very physical about Q . Indeed, let us consider henceforth the case in which the operator Q has no time dependence. It could be, for example some function of \hat{x} and \hat{p} , or for a spin-1/2 particle, the operator $|+\rangle\langle-$. Such

operator Q can easily have time-dependent expectation values, but the time dependence originates from the time dependence of the states, not from the operator Q itself.

To explore the meaning of $[H, Q]$ we begin by computing the time-derivative of the expectation value of Q :

$$\frac{d}{dt}\langle Q \rangle = \frac{d}{dt}\langle \Psi, Q\Psi \rangle = \left\langle \frac{\partial \Psi}{\partial t}, Q\Psi \right\rangle + \left\langle \Psi, Q \frac{\partial \Psi}{\partial t} \right\rangle \quad (3.37)$$

where we did not have to differentiate Q as it is time-independent. At this point we can use the Schrödinger equation to find

$$\begin{aligned} \frac{d}{dt}\langle Q \rangle &= \left\langle \frac{1}{i\hbar} H\Psi, Q\Psi \right\rangle + \left\langle \Psi, Q \frac{1}{i\hbar} H\Psi \right\rangle \\ &= \frac{i}{\hbar} \left(\langle H\Psi, Q\Psi \rangle - \langle \Psi, QH\Psi \rangle \right) \\ &= \frac{i}{\hbar} \langle \Psi, (HQ - QH)\Psi \rangle = \frac{i}{\hbar} \langle \Psi, [H, Q]\Psi \rangle \end{aligned} \quad (3.38)$$

where we used the Hermiticity of the Hamiltonian. We have thus arrived at

$$\boxed{\frac{d}{dt}\langle Q \rangle = \frac{i}{\hbar} \langle [H, Q] \rangle \quad \text{for time-independent } Q.} \quad (3.39)$$

This is a very important result. Each time you see $[H, Q]$ you should think ‘time derivative of $\langle Q \rangle$ ’. In classical mechanics one usually looks for conserved quantities, that is, functions of the dynamical variables that are time independent. In quantum mechanics a conserved operator is one whose expectation value is time independent. An operator Q is conserved if it commutes with the Hamiltonian!

With this result, the inequality (3.36) can be simplified. Indeed, using (3.39) we have

$$\left| \left\langle \frac{1}{2i}[H, Q] \right\rangle \right| = \left| \frac{1}{2i} \frac{\hbar}{i} \frac{d\langle Q \rangle}{dt} \right| = \frac{\hbar}{2} \left| \frac{d\langle Q \rangle}{dt} \right| \quad (3.40)$$

and therefore

$$\boxed{\Delta H \Delta Q \geq \frac{\hbar}{2} \left| \frac{d\langle Q \rangle}{dt} \right|, \quad \text{for time-independent } Q.} \quad (3.41)$$

This is a perfectly precise uncertainty inequality. The terms in it suggest a definition of a time Δt_Q

$$\Delta t_Q \equiv \frac{\Delta Q}{\left| \frac{d\langle Q \rangle}{dt} \right|}. \quad (3.42)$$

This quantity has units of time. It is the time it would take $\langle Q \rangle$ to change by ΔQ if both ΔQ and the velocity $\frac{d\langle Q \rangle}{dt}$ were time-independent. Since they are not necessarily so, we can view Δt_Q as the time for ‘appreciable’ change in $\langle Q \rangle$. This is certainly so when $\langle Q \rangle$ and ΔQ are roughly of the same size. In terms of Δt_Q the uncertainty inequality reads

$$\Delta H \Delta t_Q \geq \frac{\hbar}{2}. \quad (3.43)$$

This is still a precise inequality, given that Δt_Q has a concrete definition in (3.42).

As you will consider in the homework, (3.41) can be used to derive an inequality for time Δt_{\perp} that it takes for a system to become orthogonal to itself. If we call the initial state $\Psi(0)$, we call Δt_{\perp} the smallest time for which $\langle \Psi(0), \Psi(\Delta t_{\perp}) \rangle = 0$. You will be able to show that

$$\Delta H \Delta t_{\perp} \geq \frac{\hbar}{4}. \quad (3.44)$$

The speed in which a state can turn orthogonal depends on the energy uncertainty, and in quantum computation it plays a role in limiting the maximum possible speed of a computer for a fixed finite energy.

The uncertainty relation involves ΔH . It is natural to ask if this quantity is time dependent. As we show now, it is not, if the Hamiltonian is a time-independent operator. Indeed, if H is time independent, we can use H and H^2 for Q in (3.39) so that

$$\begin{aligned} \frac{d}{dt} \langle H \rangle &= \frac{i}{\hbar} \langle [H, H] \rangle = 0, \\ \frac{d}{dt} \langle H^2 \rangle &= \frac{i}{\hbar} \langle [H, H^2] \rangle = 0. \end{aligned} \quad (3.45)$$

It then follows that

$$\frac{d}{dt} (\Delta H)^2 = \frac{d}{dt} (\langle H^2 \rangle - \langle H \rangle^2) = 0. \quad (3.46)$$

showing that ΔH is a constant. So we have shown that

If H is time independent, the uncertainty ΔH is constant in time. (3.47)

The concept of conservation of energy uncertainty can be used to understand some aspects of atomic decays. Consider, for illustration the hyperfine transition in the hydrogen atom. Due to the existence of proton spin and the electron spin, the ground state of hydrogen is fourfold degenerate, corresponding to the four possible combinations of spins (up-up, up-down, down-up, down-down). The magnetic interaction between the spins actually breaks this degeneracy and produces the so-called “hyperfine” splitting. This is a very tiny split: $\delta E = 5.88 \times 10^{-6}$ ev (compare with about 13.6 ev for the ground state energy). For a hyperfine atomic transition, the emitted photon carries the energy difference δE resulting in a wavelength of 21.1cm and a frequency $\nu = 1420.405751786(30)$ MHz. The eleven significant digits of this frequency attest to the sharpness of the emission line. The issue of uncertainty arises because the excited state of the hyperfine splitting has a lifetime τ_H for decay to the ground state and emission of a photon. This lifetime is extremely long, in fact $\tau_H \sim 11$ million years ($= 3.4 \times 10^{14}$ sec, recalling that a year is about $\pi \times 10^7$ sec, accurate to better than 1%). This lifetime can be viewed as the time that takes some observable of the electron-proton system to change significantly (its total spin angular momentum, perhaps) so by the uncertainty principle it must be related to some energy uncertainty $\Delta E \sim \hbar/\tau_H \simeq 2 \times 10^{-30}$ ev. of the original excited state of the

hydrogen atom. Once the decay takes place the atom goes to the fully stable ground state, without any possible energy uncertainty. By the conservation of energy uncertainty, the photon must carry the uncertainty ΔE . But $\Delta E/\delta E \sim 3 \times 10^{-25}$, an absolutely infinitesimal effect on the photon. There is no broadening of the 21 cm line! That's one reason it is so useful in astronomy. For decays with much shorter lifetimes there can be an observable broadening of an emission line due to the energy-time uncertainty principle.

4 Lower bounds for ground state energies

You may recall that the variational principle could be used to find *upper* bounds on ground state energies. The uncertainty principle can be used to find *lower* bounds for the ground state energy of certain systems. Use below the uncertainty principle in the form $\Delta x \Delta p \geq \hbar/2$ to find rigorous lower bounds for the ground state energy of one-dimensional Hamiltonians. This is best illustrated by an example.

Consider a particle in a one-dimensional quartic potential considered earlier

$$H = \frac{p^2}{2m} + \alpha x^4, \quad (4.48)$$

where $\alpha > 0$ is a constant with units of energy over length to the fourth power. Our goal is to find a *lower bound* for the ground state energy $\langle H \rangle_{gs}$. Taking the ground state expectation value of the Hamiltonian we have

$$\langle H \rangle_{gs} = \frac{\langle p^2 \rangle_{gs}}{2m} + \alpha \langle x^4 \rangle_{gs}, \quad (4.49)$$

Recalling that

$$(\Delta p)^2 = \langle p^2 \rangle - \langle p \rangle^2, \quad (4.50)$$

we see that

$$\langle p^2 \rangle \geq (\Delta p)^2, \quad (4.51)$$

for any state of the system. We should note however, that for the ground state (or any bound state) $\langle p \rangle = 0$ so that in fact

$$\langle p^2 \rangle_{gs} = (\Delta p)_{gs}^2, \quad (4.52)$$

From the inequality $\langle A^2 \rangle \geq \langle A \rangle^2$ we have

$$\langle x^4 \rangle \geq \langle x^2 \rangle^2. \quad (4.53)$$

Moreover, just like for momentum above, $(\Delta x)^2 = \langle x^2 \rangle - \langle x \rangle^2$ leads to

$$\langle x^2 \rangle \geq (\Delta x)^2, \quad (4.54)$$

so that

$$\langle x^4 \rangle \geq (\Delta x)^4, \quad (4.55)$$

for the expectation value on arbitrary states. Therefore

$$\langle H \rangle_{gs} = \frac{\langle p^2 \rangle_{gs}}{2m} + \alpha \langle x^4 \rangle_{gs} \geq \frac{(\Delta p_{gs})^2}{2m} + \alpha (\Delta x_{gs})^4 \quad (4.56)$$

From the uncertainty principle

$$\Delta x_{gs} \Delta p_{gs} \geq \frac{\hbar}{2} \rightarrow \Delta p_{gs} \geq \frac{\hbar}{2\Delta x_{gs}}. \quad (4.57)$$

Back to the value of $\langle H \rangle_{gs}$ we get

$$\langle H \rangle_{gs} \geq \frac{\hbar^2}{8m(\Delta x_{gs})^2} + \alpha (\Delta x_{gs})^4. \quad (4.58)$$

The quantity to the right of the inequality is a function of Δx_{gs} . This function has been plotted in Figure 2.

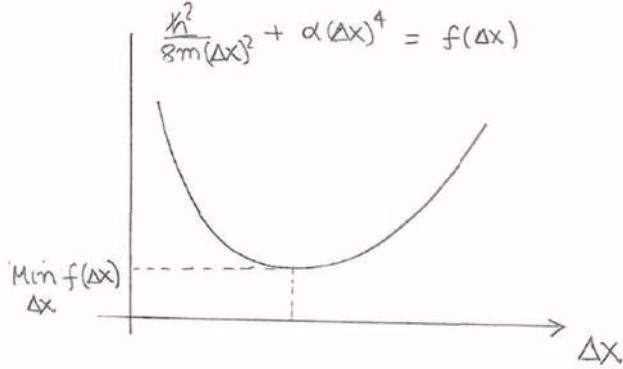


Figure 2: We have that $\langle H_{gs} \rangle \geq f(\Delta x_{gs})$ but we don't know the value of Δx_{gs} . As a result, we can only be certain that $\langle H_{gs} \rangle$ is greater than or equal to the *lowest* value the function $f(\Delta x_{gs})$ can take.

If we knew the value of Δx_{gs} we would immediately know that $\langle H \rangle_{gs}$ is bigger than the value taken by the right-hand side. This would be quite nice, since we want the highest possible lower bound. Since we don't know the value of Δx_{gs} , however, the only thing we can be sure of is that $\langle H \rangle_{gs}$ is bigger than the *lowest* value that can be taken by the expression to the right of the inequality as we vary Δx_{gs} :

$$\langle H \rangle_{gs} \geq \text{Min}_{\Delta x} \left(\frac{\hbar^2}{8m(\Delta x)^2} + \alpha (\Delta x)^4 \right). \quad (4.59)$$

The minimization problem is straightforward. In fact

$$f(x) = \frac{A}{x^2} + Bx^4 \text{ is minimized for } x^2 = 2^{-1/3} \left(\frac{A}{B} \right)^{1/3} \text{ yielding } f = 2^{1/3} \frac{3}{2} (A^2 B)^{1/3}. \quad (4.60)$$

Applied to (4.59) we obtain

$$\langle H \rangle_{gs} \geq 2^{1/3} \frac{3}{8} \left(\frac{\hbar^2 \sqrt{\alpha}}{m} \right)^{2/3} \simeq 0.4724 \left(\frac{\hbar^2 \sqrt{\alpha}}{m} \right)^{2/3}. \quad (4.61)$$

This is the final lower bound for the ground state energy. It is actually not too bad, for the ground state instead of the prefactor 0.4724, we have 0.668.

5 Diagonalization of Operators

When we have operators we wish to understand, it can be useful to find a basis on the vector space for which the operators are represented by matrices that take a simple form. Diagonal matrices are matrices where all non diagonal entries vanish. If we can find a set of basis vectors for which the matrix representing an operator is diagonal we say that the operator is **diagonalizable**.

If an operator T is diagonal in some basis (u_1, \dots, u_n) of the vector space V , its matrix takes the form $\text{diag}(\lambda_1, \dots, \lambda_n)$, with constants λ_i , and we have

$$Tu_1 = \lambda_1 u_1, \dots, Tu_n = \lambda_n u_n. \quad (5.62)$$

The basis vectors are recognized as eigenvectors with eigenvalues given by the diagonal elements. It follows that *a matrix is diagonalizable if and only if it possesses a set of eigenvectors that span the vector space*. Recall that all operators T on complex vector spaces have at least one eigenvalue and thus at least one eigenvector. But not even in complex vector spaces all operators have enough eigenvectors to span the space. Those operators cannot be diagonalized. The simplest example of such operator is provided by the two-by-two matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (5.63)$$

The only eigenvalue of this matrix is $\lambda = 0$ and the associated eigenvector is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Since a two-dimensional vector space cannot be spanned with one eigenvector, this matrix cannot be diagonalized. Having seen that the question of diagonalization of an operator is ultimately a question about its eigenvectors, we want to emphasize that the question can be formulated without referring to any basis. Bases, of course are useful, to express concretely

Suppose we have a vector space V and we have chosen a basis (v_1, \dots, v_n) such that a linear operator has a matrix representation $T_{ij}(\{v\})$ that is not diagonal. As we learned before, if we change basis to a new one (u_1, \dots, u_n) using a linear operator A such that

$$u_k = A v_k, \quad (5.64)$$

the matrix representation $T_{ij}(\{u\})$ of the operator in the new basis takes the form

$$T(\{u\}) = A^{-1} T(\{v\}) A \quad \text{or} \quad T_{ij}(\{u\}) = (A^{-1})_{ik} T_{kp}(\{v\}) A_{pj}, \quad (5.65)$$

where the matrix A_{ij} is the representation of A in the original v -basis. The operator T is diagonalizable if there is an operator A such that $T_{ij}(\{u\})$ is diagonal.

There are two pictures of the diagonalization: One can consider the operator T and state that its matrix representation is diagonal when referred to the u basis obtained by acting with A on the original v basis. Alternatively, we can view the result as the existence of a related operator $A^{-1}TA$ that is diagonal in the *original* v basis. Indeed, $Tu_i = \lambda_i u_i$ (i not summed) implies that $TA v_i = \lambda_i A v_i$

and acting with A^{-1} that $(A^{-1}TA)v_i = \lambda_i v_i$, which confirms that $A^{-1}TA$ is represented by a diagonal matrix in the original v basis. Both viewpoints are valuable.

It is useful to note that the columns of the matrix A are in fact the eigenvectors of $T(\{v\})$. We see this as follows. Since the eigenvectors are the u_k we have

$$u_k = Av_k \quad \rightarrow \quad u_k = \sum_i A_{ik} v_i . \quad (5.66)$$

Using the original basis means v_i is represented by a column vector of zeroes with a single unit entry at the i -th position. We thus find

$$u_k = \begin{pmatrix} A_{1k} \\ \vdots \\ A_{nk} \end{pmatrix} . \quad (5.67)$$

confirming that the k -th column of A is the k -th eigenvector of T .

While not all operators on complex vector spaces can be diagonalized, the situation is much improved for Hermitian operators. Recall that T is Hermitian if $T = T^\dagger$. Hermitian operators can be diagonalized, and so can unitary operators. But even more is true: the operators take diagonal form in an orthonormal basis!

An operator M is said to be **unitarily** diagonalizable if there is an *orthonormal* basis in which its matrix representation is a diagonal matrix. That basis, therefore, is an orthonormal basis of eigenvectors. Starting with an arbitrary orthonormal basis (e_1, \dots, e_n) where the matrix representation of M is $M(\{e\})$, a unitary transformation of this basis produces the orthonormal basis in which the operator takes diagonal form. More explicitly, there is a unitary matrix U ($U^\dagger = U^{-1}$) and a diagonal matrix D_M such that

$$U^\dagger M(\{e\}) U = D_M . \quad (5.68)$$

6 The Spectral Theorem

While we could prove, as most textbooks do, that Hermitian operators are unitarily diagonalizable, this result holds for a more general class of operators, called normal operators. The proof is not harder than the one for hermitian operators. An operator M is said to be **normal** if it commutes with its adjoint:

$$M \text{ is normal : } [M^\dagger, M] = 0 . \quad (6.69)$$

Hermitian operators are clearly normal. So are anti-hermitian operators ($M^\dagger = -M$ is antihermitian). Unitary operators U are normal because both $U^\dagger U$ and UU^\dagger are equal to the identity matrix and thus U and U^\dagger commute.

Exercise. If an operator M is normal show that so is $V^\dagger M V$ where V is a unitary operator.

Lemma: Let w be an eigenvector of the normal operator M : $Mw = \lambda w$. Then w is also an eigenvector of M^\dagger with complex conjugate eigenvalue:

$$M^\dagger w = \lambda^* w. \quad (6.70)$$

Proof: Define $u = (M^\dagger - \lambda^* I)w$. The result holds if u is the zero vector. To show this we compute the norm-squared of u :

$$|u|^2 = \langle u, u \rangle = \langle (M^\dagger - \lambda^* I)w, (M^\dagger - \lambda^* I)w \rangle \quad (6.71)$$

Using the adjoint property to move the operator in the first entry to the second entry:

$$|u|^2 = \langle w, (M - \lambda I)(M^\dagger - \lambda^* I)w \rangle \quad (6.72)$$

Since M and M^\dagger commute, so do the two factors in parenthesis and therefore

$$|u|^2 = \langle w, (M^\dagger - \lambda^* I)(M - \lambda I)w \rangle = 0 \quad (6.73)$$

since $(M - \lambda I)$ kills w . It follows that $u = 0$ and therefore (6.70) holds. \square

We can now state our main theorem, called the *spectral theorem*. It states that a matrix is unitarily diagonalizable if and only if it is normal. More to the point,

Spectral Theorem: Let M be an operator in a complex vector space. The vector space has a orthonormal basis comprised of eigenvectors of M if and only if M is normal.

(6.74)

Proof. It is easy to show that unitarily diagonalizable implies normality. Indeed, from (5.68) and dropping the reference to the e -basis,

$$M = UD_M U^\dagger \text{ and therefore } M^\dagger = U D_M^\dagger U^\dagger.$$

We then get

$$M^\dagger M = U D_M^\dagger D_M U^\dagger \text{ and } M M^\dagger = U D_M D_M^\dagger U^\dagger.$$

so that

$$[M^\dagger, M] = U(D_M^\dagger D_M - D_M D_M^\dagger)U^\dagger = 0,$$

because any two diagonal matrices commute.

Now let us prove that M provides a basis of orthonormal eigenvectors. The proof is by induction. The result is clearly true for $\dim V = 1$. We assume that it holds for $(n-1)$ -dimensional vector spaces and consider the case of n -dimensional V . Let M be an $n \times n$ matrix referred to the orthonormal basis $(|1\rangle, \dots, |n\rangle)$ of V so that $M_{ij} = \langle i|M|j \rangle$. We know there is at least one eigenvalue λ_1 with a non-zero eigenvector $|x_1\rangle$ of unit norm:

$$M|x_1\rangle = \lambda_1|x_1\rangle \text{ and } M^\dagger|x_1\rangle = \lambda_1^*|x_1\rangle, \quad (6.75)$$

in view of the Lemma. There is, we claim, a unitary matrix U_1 such that

$$|x_1\rangle = U_1|1\rangle \rightarrow U_1^\dagger|x_1\rangle = |1\rangle. \quad (6.76)$$

U_1 is not unique and can be constructed as follows: extend $|x_1\rangle$ to an orthonormal basis $|x_1\rangle, \dots, |x_N\rangle$ using Gram-Schmidt. Then write $U_1 = \sum_i |x_i\rangle\langle i|$. Define now

$$M_1 \equiv U_1^\dagger M U_1. \quad (6.77)$$

M_1 is also normal and $M_1|1\rangle = U_1^\dagger M U_1|1\rangle = U_1^\dagger M|x_1\rangle = \lambda_1 U_1^\dagger|x_1\rangle = \lambda_1|1\rangle$, so that

$$M_1|1\rangle = \lambda_1|1\rangle. \quad (6.78)$$

Let us now examine the explicit form of the matrix M_1 :

$$\langle j|M_1|1\rangle = \lambda_1\langle j|1\rangle = \lambda_1\delta_{i,j}, \quad (6.79)$$

which says that the first column of M_1 has zeroes in all entries except the first. Moreover

$$\langle 1|M_1|j\rangle = (\langle j|M_1^\dagger|1\rangle)^* = (\lambda_1^*\langle j|1\rangle)^* = \lambda_1\langle 1|j\rangle = \lambda_1\delta_{i,j}, \quad (6.80)$$

where we used $M_1^\dagger|1\rangle = \lambda_1^*|1\rangle$ which follows from the normality of M_1 . It follows from the two last equations that M_1 , in the original basis, takes the form

$$M_1 = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & M' & \\ 0 & & & \end{pmatrix}.$$

Since M_1 is normal, one can see that M' is a normal $(n-1)$ -by- $(n-1)$ matrix. By the induction hypothesis M' can be unitarily diagonalized so that $U'^\dagger M' U'$ is diagonal for some $(n-1)$ -by- $(n-1)$ unitary matrix U' . The matrix U' can be extended to an n -by- n unitary matrix \hat{U} as follows

$$\hat{U} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & U' & \\ 0 & & & \end{pmatrix}. \quad (6.81)$$

It follows that $\hat{U}^\dagger M_1 \hat{U} = \hat{U}^\dagger U_1^\dagger M U_1 \hat{U} = (U_1 \hat{U})^\dagger M (U_1 \hat{U})$ is diagonal, proving the desired result. \square .

Of course this theorem implies that Hermitian and unitary operators are unitarily diagonalizable. In other words the eigenvectors form an orthonormal basis. This is true whether or not there are degeneracies in the spectrum. The proof does not require discussion of this as a special case. If an eigenvalue of M is degenerate and appears k times, then there are k orthonormal eigenvectors associated with the corresponding k -dimensional M -invariant subspace of the vector space.

- - - - -

We conclude this section with a description of the general situation that we may encounter when diagonalizing a normal operator T . In general, we expect degeneracies in the eigenvalues so that each eigenvalue λ_k is repeated $d_k \geq 1$ times. An eigenvalue λ_k is degenerate if $d_k > 1$. It follows that V has T -invariant subspaces of different dimensionalities. Let U_k denote the T -invariant subspace of dimension $d_k \geq 1$ spanned by eigenvectors with eigenvalue λ_k :

$$U_k \equiv \{v \in V \mid T v = \lambda_k v\}, \quad \dim U_k = d_k. \quad (6.82)$$

By the spectral theorem U_k has a basis comprised by d_k orthonormal eigenvectors $(u_1^{(k)}, \dots, u_{d_k}^{(k)})$. Note that while the addition of eigenvectors with different eigenvalues does not give eigenvectors, in the subspace U_k all vectors are eigenvectors with the same eigenvalue, and that's why addition makes sense, U_k as defined is a vector space, and adding eigenvectors in U_k gives eigenvectors. The full space V is decomposed as the direct sum of the invariant subspaces of T :

$$V = U_1 \oplus U_2 \oplus \dots \oplus U_m, \quad \dim V = \sum_{i=1}^m d_i, \quad m \geq 1. \quad (6.83)$$

All U_i subspaces are guaranteed to be orthogonal to each other. In fact the full list of eigenvectors is a list of orthonormal vectors that form a basis for V is conveniently ordered as follows:

$$(u_1^{(1)}, \dots, u_{d_1}^{(1)}, \dots, u_1^{(m)}, \dots, u_{d_m}^{(m)}). \quad (6.84)$$

The matrix T is manifestly diagonal in this basis because each vector above is an eigenvector of T and is orthogonal to all others. The matrix representation of T reads

$$T = \text{diag} \left(\underbrace{\lambda_1, \dots, \lambda_1}_{d_1 \text{ times}}, \dots, \underbrace{\lambda_m, \dots, \lambda_m}_{d_m \text{ times}} \right) \quad (6.85)$$

This is clear because the first d_1 vectors in the list are in U_1 , the second d_2 vectors are in U_2 , and so on and so forth until the last d_m vectors are in U_m .

If we had no degeneracies in the spectrum the basis (6.84) (with $d_i = 1$ for all i) would be rather unique if we require the matrix representation of T to be unchanged. Each vector could be multiplied by a phase. On the other hand, with degeneracies that the list (6.84) can be changed considerably without changing the matrix representation of T . Let V_k be a unitary operator on U_k , for each $k = 1, \dots, m$. We claim that the following basis of eigenvectors leads to the same matrix T :

$$(V_1 u_1^{(1)}, \dots, V_1 u_{d_1}^{(1)}, \dots, V_m u_1^{(m)}, \dots, V_m u_{d_m}^{(m)}). \quad (6.86)$$

Indeed, this is still a collection of eigenvectors of T with each of them orthogonal to the rest. Moreover, the first d_1 vectors are in U_1 , the second d_2 vectors are in U_2 and so on and so forth. More explicitly, for example, within U_k

$$\langle V_k u_i^{(k)}, T(V_k u_j^{(k)}) \rangle = \lambda_k \langle V_k u_i^{(k)}, V_k u_j^{(k)} \rangle = \lambda_k \langle u_i^{(k)}, u_j^{(k)} \rangle = \lambda_k \delta_{ij} \quad (6.87)$$

showing that in the U_k subspace the matrix for T is still diagonal with all entries equal to λ_k .

7 Simultaneous Diagonalization of Hermitian Operators

We say that two operators S and T in a vector space V operators can be **simultaneously diagonalized** if there is some basis of V in which both the matrix representation of S and the matrix representation of T are diagonal. It then follows that each vector in this basis is an eigenvector of S and an eigenvector of T .

A necessary condition for simultaneous diagonalization is that the operators S and T commute. Indeed, if they can be simultaneously diagonalized there is a basis where both are diagonal and they manifestly commute. If the operators don't commute, this is a basis-independent statement and therefore a simultaneous diagonal presentation cannot exist. Since arbitrary linear operators S and T on a complex vector space cannot be diagonalized, the vanishing of $[S, T]$ does not guarantee simultaneous diagonalization. But if the operators are Hermitian it does, as we show now.

Theorem. If S and T are commuting Hermitian operators they can be simultaneously diagonalized.

Proof. The main complication is that degeneracies in the spectrum require an some discussion. Either both operators have degeneracies or one has no degeneracies. Without loss of generality we can assume that there are two cases to consider

- (i) There is no degeneracy in the spectrum of T or,
- (ii) Both T and S have degeneracies in their spectrum.

Consider case (i) first. Since T is non-degenerate there is a basis (u_1, \dots, u_n) of eigenvectors of T with different eigenvalues

$$Tu_i = \lambda_i u_i, \quad i \text{ not summed}, \quad \lambda_i \neq \lambda_j \text{ for } i \neq j. \quad (7.88)$$

We now want to understand what kind of vector is Su_i . For this we act with T on it

$$T(Su_i) = S(Tu_i) = S(\lambda_i u_i) = \lambda_i (Su_i), \quad (7.89)$$

It follows that Su_i is also an eigenvector of T with eigenvalue λ_i , thus it must equal u_i , up to scale,

$$Su_i = \omega_i u_i, \quad (7.90)$$

showing that u_i is also an eigenvector of S , this time with eigenvalue ω_i . Thus any eigenvector of T is also an eigenvector of S , showing that these operators are simultaneously diagonalizable.

Now consider case (ii). Since T has degeneracies, as explained in the previous section, we have a decomposition of V in T -invariant subspaces U_k spanned by eigenvectors:

$$\begin{aligned} U_k &\equiv \{u \mid Tu = \lambda_k u\}, \quad \dim U_k = d_k \quad V = U_1 \oplus \dots \oplus U_m, \\ \text{orthonormal basis for } V : \quad &(u_1^{(1)}, \dots, u_{d_1}^{(1)}, \dots, u_1^{(m)}, \dots, u_{d_m}^{(m)}). \\ T &= \text{diag} \left(\underbrace{\lambda_1, \dots, \lambda_1}_{d_1 \text{ times}}, \dots, \underbrace{\lambda_m, \dots, \lambda_m}_{d_m \text{ times}} \right) \text{ in this basis.} \end{aligned} \quad (7.91)$$

We also explained that the alternative orthonormal basis of V

$$(V_1 u_1^{(1)}, \dots V_1 u_{d_1}^{(1)}, \dots, V_m u_1^{(m)}, \dots, V_m u_{d_m}^{(m)}). \quad (7.92)$$

leads to the same matrix for T when each V_k is a unitary operator on U_k .

We now claim that the U_k are also S -invariant subspaces! To show this let $u \in U_k$ and examine the vector Su . We have

$$T(Su) = S(Tu) = \lambda_k Su \rightarrow Su \in U_k. \quad (7.93)$$

We use the subspaces U_k and the basis (7.91) to organize the matrix representation of S in blocks. It follows that this matrix must have *block-diagonal* form since each subspace is S -invariant and orthogonal to all other subspaces. We cannot guarantee, however, that S is diagonal within each square block because $Su_i^{(k)} \in U_k$ but we have no reason to believe that $Su_i^{(k)}$ points along $u_i^{(k)}$.

Since S restricted to each S -invariant subspace U_k is hermitian we can find an orthonormal basis of U_k in which the matrix S is diagonal. This new basis is unitarily related to the original basis $(u_1^{(k)}, \dots, u_{d_k}^{(k)})$ and thus takes the form $(V_k u_1^{(k)}, \dots, V_k u_{d_k}^{(k)})$ with V_k a unitary operator in U_k . Note that the eigenvalues of S in this block need not be degenerate. Doing this for each block, we find a basis of the form (7.92) in which S is diagonal. But T is still diagonal in this new basis, so both S and T have been simultaneously diagonalized. \square

Remarks:

1. Note that the above proof gives an algorithmic way to produce the common list of eigenvectors. One diagonalizes one of the matrices and constructs the second matrix in the basis of eigenvectors of the first. These second matrix is block diagonal, where the blocks are organized by the degeneracies in the spectrum of the first matrix. One must then diagonalize within the blocks and is guaranteed that the new basis that works for the second matrix also works for the first.
2. If we had to simultaneously diagonalize three different commuting Hermitian operators S_1, S_2 and S_3 , all of which have degenerate spectra, we would proceed as follows. We diagonalize S_1 and fix a basis in which S_1 is diagonal. In this basis we must find that S_2 and S_3 have exactly the same block structure. The corresponding block matrices are simply the matrix representations of S_2 and S_3 in each of the invariant spaces U_k appearing in the diagonalization of S_1 . Since S_2 and S_3 commute, their restrictions to U_k commute. These restrictions can be diagonalized simultaneously, as guaranteed by our theorem which works for two matrices. The new basis in U_k that makes the restriction of S_2 and S_3 diagonal, will not disturb the diagonal form of S_1 in this block. This is repeated for each block, until we get a common basis of eigenvectors.
3. An inductive algorithm is clear. If we know how to simultaneously diagonalize n commuting Hermitian operators we can diagonalize $n + 1$ of them, call them S_1, \dots, S_{n+1} , as follows. We diagonalize S_1 and then consider the remaining n operators in the basis that makes S_1 diagonal.

We are guaranteed a common block structure for the n operators. The problem becomes one of simultaneous diagonalization of n commuting Hermitian block matrices, which is assumed known by the induction argument.

Corollary. If $\{S_1, \dots, S_n\}$ is a set of mutually commuting Hermitian operators they can all be simultaneously diagonalized.

8 Complete Set of Commuting Observables

We have discussed the problem of finding eigenvectors and eigenvalues of a Hermitian operator S . This hermitian operator is thought as a quantum mechanical observable. The eigenvectors of S are physical states of the system in which the observable S can be measured without uncertainty. The result of the measurement is the eigenvalue associated with the eigenvector.

If the Hermitian operator S has a non-degenerate spectrum, all eigenvalues are different and we have a rather nice situation in which each eigenvector can be uniquely distinguished by labeling it with the corresponding eigenvalue of S . The physical quantity associated with the observable can be used to distinguish the various eigenstates. Moreover, these eigenstates provide an orthonormal basis for the full vector space. In this case the operator S provides a “complete set of commuting observables” or a CSCO, in short. The set here has just one observable, the operator S .

The situation is more nontrivial if the Hermitian operator S exhibits degeneracies in its spectrum. This means that V has an S -invariant subspace of dimension $d > 1$, spanned by orthonormal eigenvectors (u_1, \dots, u_d) all of which have S eigenvalue λ . This time, the eigenvalue of S does not allow us to distinguish or to label uniquely the basis eigenstates of the invariant subspace. Physically this is a deficient situation, as we have explicitly different states – the various u_i ’s – that we can’t tell apart by the measurement of S alone. This time S does not provide a CSCO. Labeling eigenstates by the S eigenvalue does not suffice to distinguish them.

We are thus physically motivated to find another Hermitian operator T that is compatible with S . Two Hermitian operators are said to be **compatible observables** if they commute, since then we can find a basis of V comprised by simultaneous eigenvectors of the operators. These states can be labeled by two observables, namely, the two eigenvalues. If we are lucky, the basis eigenstates in each of the S -invariant subspaces of dimension higher than one can be organized into T eigenstates of different eigenvalues. In this case T breaks the spectral degeneracy of S and using T eigenvalues as well as S eigenvalues we can label uniquely a basis of orthonormal states of V . In this case we say that S and T form a CSCO.

We have now given enough motivation for a definition of a complete set of commuting observables. Consider a set of commuting observables, namely, a set $\{S_1, \dots, S_k\}$ of Hermitian operators acting on a complex vector space V that represents the physical state-space of some quantum system. By the theorem in the previous section, we can find an orthonormal basis of vectors in V such that each vector

is an eigenstate of every operator in the set. Assume that each eigenstate in the basis is labeled by the eigenvalues of the S_i operators. The set $\{S_1, \dots, S_k\}$ is said to be a **complete set of commuting observables** if no two states have the same labels.

It is a physically motivated assumption that for any physical quantum system there is a complete set of commuting observables, for otherwise there is no physical way to distinguish the various states that span the vector space. So in any physical problem we are urged to find such complete set, and we must include operators in such set until all degeneracies are broken. A CSCO need not be unique. Once we have a complete set of commuting observables, adding another observable causes no harm, although it is not necessary. Also, if (S_1, S_2) form a CSCO, so will $(S_1 + S_2, S_1 - S_2)$. Ideally, we want the smallest set of operators.

The first operator that is usually included in a CSCO is the Hamiltonian H . For bound state problems in one dimension, energy eigenstates are non-degenerate and thus the energy can be used to label uniquely the H -eigenstates. A simple example is the infinite square well. Another example is the one-dimensional harmonic oscillator. In such cases H forms the CSCO. If we have, however, a two-dimensional isotropic harmonic oscillator in the (x, y) plane, the Hamiltonian has degeneracies. At the first excited level we can have the first excited state of the x harmonic oscillator or, at the same energy, the first excited state of the y harmonic oscillator. We thus need another observable that can be used to distinguish these states. There are several options, as you will discuss in the homework.

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 10: Solving the Time-Independent Schrödinger Equation

B. Zwiebach
March 14, 2016

Contents

1 Stationary States	1
2 Solving for Energy Eigenstates	3
3 Free particle on a circle.	6

1 Stationary States

Consider the Schrödinger equation for the wavefunction $\Psi(x, t)$ with the assumption that the potential energy V is time independent:

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi(x, t) = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right) \Psi(x, t), \quad (1.1)$$

where we displayed the form of the Hamiltonian operator \hat{H} with the time independent potential $V(x)$. Stationary states are a very useful class of solutions of this differential equation. The signature property of a stationary state is that the position and the time dependence of the wavefunction factorize. Namely,

$$\Psi(x, t) = g(t)\psi(x), \quad (1.2)$$

for some functions g and ψ . For such a *separable* solution to exist we need the potential to be time independent, as we will see below. The solution $\Psi(x, t)$ is time-dependent but it is called stationary because of a property of observables. The expectation value of observables with no explicit time dependence in arbitrary states has time dependence. On a stationary state they do not have time dependence, as we will demonstrate.

Let us use the ansatz (1.2) for Ψ in the Schrödinger equation. We then find

$$\left(i\hbar \frac{dg(t)}{dt} \right) \psi(x) = g(t) \hat{H}\psi(x), \quad (1.3)$$

because $g(t)$ can be moved across \hat{H} . We can then divide this equation by $\Psi(x, t) = g(t)\psi(x)$, giving

$$i\hbar \frac{1}{g(t)} \frac{dg(t)}{dt} = \frac{1}{\psi(x)} \hat{H}\psi(x). \quad (1.4)$$

The left side is a function of only t , while the right side is a function of only x (a time dependent potential would have spoiled this). The only way the two sides can equal each other for all values of t and x is for both sides to be equal to a *constant E* with units of energy because \hat{H} has units of energy. We therefore get two separate equations. The first reads

$$i\hbar \frac{dg}{dt} = Eg. \quad (1.5)$$

This is solved by

$$g(t) = e^{-iEt/\hbar}, \quad (1.6)$$

and the most general solution is simply a constant times the above right-hand side. From the x -dependent side of the equality we get

$$\hat{H}\psi(x) = E\psi(x). \quad (1.7)$$

This equation is an eigenvalue equation for the Hermitian operator \hat{H} . We showed that the eigenvalues of Hermitian operators must be real, thus the constant E **must be real**. The equation above is called the **time-independent Schrödinger equation**. More explicitly it reads

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right) \psi(x) = E\psi(x), \quad (1.8)$$

Note that this equation does not determine the overall normalization of ψ . Therefore we can write the full solution without loss of generality using the $g(t)$ given above:

$$\text{Stationary state: } \Psi(x, t) = e^{-iEt/\hbar} \psi(x), \quad \text{with } E \in \mathbb{R} \text{ and } \hat{H}\psi = E\psi. \quad (1.9)$$

Note that not only is $\psi(x)$ an eigenstate of the Hamiltonian operator \hat{H} , the full stationary state is also an \hat{H} eigenstate

$$\hat{H}\Psi(x, t) = E\Psi(x, t), \quad (1.10)$$

since the time dependent function in Ψ cancels out.

We have noted that the energy E must be real. If it was not we would also have trouble normalizing the stationary state consistently. The normalization condition for Ψ , if E is not real, would give

$$\begin{aligned} 1 &= \int dx \Psi^*(x, t)\Psi(x, t) = \int dx e^{iE^*t/\hbar} e^{-iEt/\hbar} \psi^*(x)\psi(x) \\ &= e^{i(E^*-E)t/\hbar} \int dx \psi^*(x)\psi(x) = e^{2\text{Im}(E)t/\hbar} \int dx \psi^*(x)\psi(x). \end{aligned} \quad (1.11)$$

The final expression has a time dependence due to the exponential. On the other hand the normalization condition states that this expression must be equal to one. It follows that the exponent must be zero, i.e., E is real. Given this, we also see that the normalization condition yields

$$\int_{-\infty}^{\infty} dx \psi^*(x)\psi(x) = 1. \quad (1.12)$$

How do we interpret the eigenvalue E ? Using (1.10) we see that the expectation value of \hat{H} on the state Ψ is indeed the energy

$$\langle \hat{H} \rangle_{\Psi} = \int dx \Psi^*(x, t)\hat{H}\Psi(x, t) = \int dx \Psi^*(x, t)E\Psi(x, t) = E \int dx \Psi^*(x, t)\Psi(x, t) = E, \quad (1.13)$$

Since the stationary state is an eigenstate of \hat{H} , the uncertainty $\Delta\hat{H}$ of the Hamiltonian in a stationary state is zero.

There are two important observations on stationary states:

- (1) The expectation value of any time-independent operator \hat{Q} on a stationary state Ψ is time-independent:

$$\begin{aligned}\langle Q \rangle_{\Psi(x,t)} &= \int dx \Psi^*(x,t) \hat{Q} \Psi(x,t) = \int dx e^{iEt/\hbar} \psi^*(x) \hat{Q} e^{-iEt/\hbar} \psi(x) \\ &= \int dx e^{iEt/\hbar} e^{-iEt/\hbar} \psi^*(x) \hat{Q} \psi(x) = \int dx \psi^*(x) \hat{Q} \psi(x) = \langle Q \rangle_{\psi(x)},\end{aligned}\quad (1.14)$$

since the last expectation value is manifestly time independent.

- (2) The superposition of stationary states with different energies not stationary. This is clear because a stationary state requires a factorized solution of the Schrödinger equation: if we add two factorized solutions with different energies they will have different time dependence and the total state cannot be factorized. We now show that that a time-independent observable \hat{Q} may have a time-dependent expectation values in such a state. Consider a superposition

$$\Psi(x,t) = c_1 e^{-iE_1 t/\hbar} \psi_1(x) + c_2 e^{-iE_2 t/\hbar} \psi_2(x), \quad (1.15)$$

where ψ_1 and ψ_2 are \hat{H} eigenstates with energies E_1 and E_2 , respectively. Consider a Hermitian operator \hat{Q} . With the system in state (1.15), its expectation value is

$$\begin{aligned}\langle Q \rangle_{\Psi} &= \int_{-\infty}^{\infty} dx \Psi^*(x,t) \hat{Q} \Psi(x,t) \\ &= \int_{-\infty}^{\infty} dx (c_1^* e^{iE_1 t/\hbar} \psi_1^*(x) + c_2^* e^{iE_2 t/\hbar} \psi_2^*(x)) (c_1 e^{-iE_1 t/\hbar} \hat{Q} \psi_1(x) + c_2 e^{-iE_2 t/\hbar} \hat{Q} \psi_2(x)) \\ &= \int_{-\infty}^{\infty} dx (|c_1|^2 \psi_1^* \hat{Q} \psi_1 + |c_2|^2 \psi_2^* \hat{Q} \psi_2 + c_1^* c_2 e^{i(E_1 - E_2)t/\hbar} \psi_1^* \hat{Q} \psi_2 + c_2^* c_1 e^{-i(E_1 - E_2)t/\hbar} \psi_2^* \hat{Q} \psi_1)\end{aligned}\quad (1.16)$$

We now see the possible time dependence arising from the cross terms. The first two terms are simple time-independent expectation values. Using the hermitically of \hat{Q} in the last term we then get

$$\begin{aligned}\langle Q \rangle_{\Psi} &= |c_1|^2 \langle Q \rangle_{\psi_1} + |c_2|^2 \langle Q \rangle_{\psi_2} \\ &\quad + c_1^* c_2 e^{i(E_1 - E_2)t/\hbar} \int_{-\infty}^{\infty} dx \psi_1^* \hat{Q} \psi_2 + c_1 c_2^* e^{-i(E_1 - E_2)t/\hbar} \int_{-\infty}^{\infty} dx \psi_1 (\hat{Q} \psi_2)^*\end{aligned}\quad (1.17)$$

The last two terms are complex conjugates of each other and therefore

$$\langle Q \rangle_{\Psi} = |c_1|^2 \langle Q \rangle_{\psi_1} + |c_2|^2 \langle Q \rangle_{\psi_2} + 2 \operatorname{Re} \left[c_1^* c_2 e^{i(E_1 - E_2)t/\hbar} \int_{-\infty}^{\infty} dx \psi_1^* \hat{Q} \psi_2 \right]. \quad (1.18)$$

We see that this expectation value is time-dependent if $E_1 \neq E_2$ and $(\psi_1, Q \psi_2)$ is nonzero. The full expectation value $\langle Q \rangle_{\Psi}$ is real, as it must be for any Hermitian operator.

2 Solving for Energy Eigenstates

We will now study solutions to the time-independent Schrödinger equation

$$\hat{H} \psi(x) = E \psi(x). \quad (2.19)$$

For a given Hamiltonian \hat{H} we are interested in finding the eigenstates ψ and the eigenvalues E , which happen to be the corresponding energies. Perhaps the most interesting feature of the above equation is that generally the value of E cannot be arbitrary. Just like finite size matrices have a set of eigenvalues, the above, time-independent Schrödinger equation may have a discrete set of possible energies. A continuous set of possible energies is also allowed and sometimes important. There are indeed many solutions for any given potential. Assuming for convenience that the eigenstates and their energies can be counted we write

$$\begin{aligned} \psi_1(x), & \quad E_1 \\ \psi_2(x), & \quad E_2 \\ \vdots & \quad \vdots \end{aligned} \tag{2.20}$$

Our earlier discussion of Hermitian operators applies here. The energy eigenstates can be organized to form a *complete set of orthonormal functions*:

$$\int \psi_i^*(x) \psi_j(x) = \delta_{ij}. \tag{2.21}$$

Consider the time-independent Schrödinger equation written as

$$\frac{d^2\psi}{dx^2} = -\frac{2m}{\hbar^2}(E - V(x))\psi. \tag{2.22}$$

The solutions $\psi(x)$ depend on the properties of the potential $V(x)$. It is hard to make general statements about the wavefunction unless we restrict the types of potentials. We will certainly consider continuous potentials. We also consider potentials that are not continuous but are piece-wise continuous, that is, they have a number of discontinuities. Our potentials can easily fail to be bounded. We allow delta functions in one-dimensional potentials but do not consider powers or derivatives of delta functions. We allow for potentials that become plus infinity beyond certain points. These points represent hard walls.

We want to understand general properties of ψ and the behavior of ψ at points where the potential $V(x)$ may have discontinuities or other singularities. We claim: **we must have a continuous wavefunction**. If ψ is discontinuous then ψ' contains delta-functions and ψ'' in the above left-hand side contains derivatives of delta functions. This would require the right-hand side to have derivatives of delta functions, and those would have to appear in the potential. Since we have declared that our potentials contain no derivatives of delta functions we must indeed have a continuous ψ .

Consider now four possibilities concerning the potential:

- (1) $V(x)$ is continuous. In this case the continuity of $\psi(x)$ and (2.22) imply ψ'' is also continuous. This requires ψ' continuous.
- (2) $V(x)$ has finite discontinuities. In this case ψ'' has finite discontinuities: it includes the product of a continuous ψ against a discontinuous V . But then ψ' must be continuous, with non-continuous derivative.
- (3) $V(x)$ contains delta functions. In this case ψ'' also contains delta functions: it is proportional to the product of a continuous ψ and a delta function in V . Thus ψ' has finite discontinuities.

(4) $V(x)$ contains a hard wall. A potential that is finite immediately to the left of $x = a$ and becomes infinite for $x > a$ is said to have a hard wall at $x = a$. In such a case, the wavefunction will vanish for $x \geq a$. The slope ψ' will be finite as $x \rightarrow a$ from the left, and will vanish for $x > a$. Thus ψ' is discontinuous at the wall.

In the first two cases ψ' is continuous, and in the second two it can have a finite discontinuity. In conclusion

Both ψ and ψ' are continuous unless the potential has delta functions or hard walls in which cases ψ' may have finite discontinuities.

Let us give an slightly different argument for the continuity of ψ and $\frac{d\psi}{dx}$ in the case of a potential with a finite discontinuity, such as the step shown in Fig. 1.

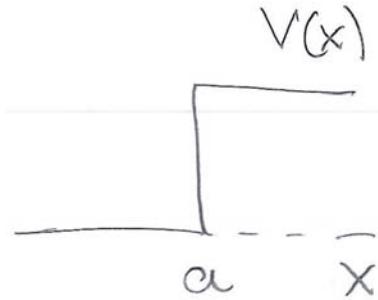


Figure 1: A potential $V(x)$ with a finite discontinuity at $x = a$.

Integrate both sides of (2.22) $a - \epsilon$ to $a + \epsilon$, and then take $\epsilon \rightarrow 0$. We find

$$\int_{a-\epsilon}^{a+\epsilon} dx \frac{d}{dx} \left(\frac{d\psi}{dx} \right) = -\frac{2m}{\hbar^2} \int_{a-\epsilon}^{a+\epsilon} dx (E - V(x))\psi(x). \quad (2.24)$$

The left-hand side integrand is a total derivative so we have

$$\frac{d\psi}{dx} \Big|_{a+\epsilon} - \frac{d\psi}{dx} \Big|_{a-\epsilon} = \frac{2m}{\hbar^2} \int_{a-\epsilon}^{a+\epsilon} dx (V(x) - E)\psi(x). \quad (2.25)$$

By definition, the discontinuity in the derivative of ψ at $x = a$ is the limit as $\epsilon \rightarrow 0$ of the left-hand side:

$$\Delta_a \left(\frac{d\psi}{dx} \right) \equiv \lim_{\epsilon \rightarrow 0} \left(\frac{d\psi}{dx} \Big|_{a+\epsilon} - \frac{d\psi}{dx} \Big|_{a-\epsilon} \right). \quad (2.26)$$

Back in (2.25) we then have

$$\Delta_a \left(\frac{d\psi}{dx} \right) = \lim_{\epsilon \rightarrow 0} \frac{2m}{\hbar^2} \int_{a-\epsilon}^{a+\epsilon} dx (V(x) - E)\psi(x). \quad (2.27)$$

The potential V is discontinuous but not infinite around $x = a$, nor is ψ infinite around $x = a$ and, of course, E is assumed finite. As the integral range becomes vanishingly small about $x = a$ the integrand remains finite and the integral goes to zero. We thus have

$$\Delta_a \left(\frac{d\psi}{dx} \right) = 0. \quad (2.28)$$

There is no discontinuity in $\frac{d\psi}{dx}$. This gives us one of our boundary conditions.

To learn about the continuity of ψ we reconsider the first integral of the differential equation. The integration that led to (2.25) now applied to the range from $x_0 < a$ to x yields

$$\frac{d\psi(x)}{dx} = \left. \frac{d\psi}{dx} \right|_{x_0} - \frac{2m}{\hbar} \int_{x_0}^x (E - V(x')) dx'. \quad (2.29)$$

Note that the integral on the right is a bounded function of x . We now integrate again from $a - \epsilon$ to $a + \epsilon$. Since the first term on the right-hand side is a constant we find

$$\psi(a + \epsilon) - \psi(a - \epsilon) = 2\epsilon \left. \frac{d\psi}{dx} \right|_{x_0} - \frac{2m}{\hbar} \int_{a-\epsilon}^{a+\epsilon} dx \int_{x_0}^x (E - V(x')). \quad (2.30)$$

Taking the $\epsilon \rightarrow 0$ limit, the first term on the right-hand side clearly vanishes and the second term goes to zero because $\int_{x_0}^x dx' (E - V(x'))$ is a bounded function of x . As a result we have

$$\Delta_a \psi = 0, \quad (2.31)$$

showing that the wavefunction is continuous at $x = a$. This is our second boundary condition.

3 Free particle on a circle.

Consider now the problem of a particle confined to a circle of circumference L . The coordinate along the circle is called x and we can view the circle as the interval $x \in [0, L]$ with the endpoints identified. It is perhaps clearer mathematically to think of the circle as the full real line x with the identification

$$x \sim x + L, \quad (3.1)$$

which means that two points whose coordinates are related in this way are to be considered **the same point**. It follows that we have the periodicity condition

$$\psi(x + L) = \psi(x). \quad (3.2)$$

From this it follows that not only ψ is periodic but all of its derivatives are also periodic.

The particle is assumed to be free and therefore $V(x) = 0$. The time-independent Schrödinger equation is then

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} = E \psi(x). \quad (3.3)$$

Before we solve this, let us show that any solution must have $E \geq 0$. For this multiply the above equation by $\psi^*(x)$ and integrate over the circle $x \in [0, L]$. Since ψ is normalized we get

$$-\frac{\hbar^2}{2m} \int_0^L \psi^*(x) \frac{d^2\psi}{dx^2} dx = E \int \psi^*(x) \psi(x) dx = E. \quad (3.4)$$

The integrand on the left hand side can be rewritten as

$$-\frac{\hbar^2}{2m} \int_0^L \left[\frac{d}{dx} \left(\psi^* \frac{d\psi}{dx} \right) - \frac{d\psi^*}{dx} \frac{d\psi}{dx} \right] dx = E. \quad (3.5)$$

and the total derivative can be integrated

$$-\frac{\hbar^2}{2m} \left[\left(\psi^* \frac{d\psi}{dx} \right) \Big|_{x=L} - \left(\psi^* \frac{d\psi}{dx} \right) \Big|_{x=0} \right] + \frac{\hbar^2}{2m} \int_0^L \left| \frac{d\psi}{dx} \right|^2 dx = E. \quad (3.6)$$

Since $\psi(x)$ and its derivatives are periodic, the contributions from $x = L$ and $x = 0$ cancel out and we are left with

$$E = \frac{\hbar^2}{2m} \int_0^L \left| \frac{d\psi}{dx} \right|^2 dx \geq 0, \quad (3.7)$$

which establishes our claim. We also see that $E = 0$ requires ψ constant (and nonzero!).

Having shown that all solutions must have $E \geq 0$ let us go back to the Schrödinger equation, which can be rewritten as

$$\frac{d^2\psi}{dx^2} = -\frac{2mE}{\hbar^2} \psi. \quad (3.8)$$

We can then define k via

$$k^2 \equiv \frac{2mE}{\hbar^2} \geq 0. \quad (3.9)$$

Since $E \geq 0$, the constant k is real. Note that this definition is very natural, since it makes

$$E = \frac{\hbar^2 k^2}{2m}, \quad (3.10)$$

which means that, as usual, $p = \hbar k$. Using k^2 the differential equation becomes the familiar

$$\frac{d^2\psi}{dx^2} = -k^2 \psi. \quad (3.11)$$

We could write the general solution in terms of sines and cosines of kx , but let's use complex exponentials:

$$\psi(x) \sim e^{ikx}. \quad (3.12)$$

This solves the differential equation and, moreover, it is a momentum eigenstate. The periodicity condition (3.2) requires

$$e^{ik(x+L)} = e^{ikx} \rightarrow e^{ikL} = 1 \rightarrow kL = 2\pi n, \quad n \in \mathbb{Z}. \quad (3.13)$$

We see that momentum is quantized because the wavenumber is quantized! The wavenumber has discrete possible values

$$k_n \equiv \frac{2\pi n}{L}, \quad n \in \mathbb{Z}. \quad (3.14)$$

All integers positive and negative are allowed and are in fact necessary because they all correspond to *different* values of the momentum $p_n = \hbar k_n$. The solutions to the Schrödinger equation can then be indexed by the integer n :

$$\psi_n(x) = N e^{ik_n x}, \quad (3.15)$$

where N is a real normalization constant. Its value is determined from

$$1 = \int_0^L \psi_n^*(x) \psi_n(x) dx = \int_0^L N^2 dx = N^2 L \rightarrow N = \frac{1}{\sqrt{L}}, \quad (3.16)$$

so we have

$$\psi_n(x) = \frac{1}{\sqrt{L}} e^{ik_n x} = \frac{1}{\sqrt{L}} e^{\frac{2\pi i n x}{L}}. \quad (3.17)$$

The associated energies are

$$E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2 4\pi^2 n^2}{2m L^2} = \frac{2\pi^2 \hbar^2 n^2}{m L^2}. \quad (3.18)$$

There are infinitely many energy eigenstates. We have degenerate states because E_n is just a function of $|n|$ and thus the same for n and $-n$. Indeed ψ_n and ψ_{-n} both have energy E_n . The only nondegenerate eigenstate is $\psi_0 = \frac{1}{\sqrt{L}}$, which is a constant wavefunction with zero energy.

Whenever we find degenerate energy eigenstates we must wonder what makes those states different, given that they have the same energy. To answer this one must find an observable that takes different values on the states. Happily, in our case we know the answer. Our degenerate states can be distinguished by their momentum: ψ_n has momentum $2\pi n \frac{\hbar}{L}$ and ψ_{-n} has momentum $(-2\pi n \frac{\hbar}{L})$.

Given two degenerate energy eigenstates, any linear combination of these states is an eigenstate with the same energy. Indeed if

$$\hat{H}\psi_1 = E\psi_1, \quad \hat{H}\psi_2 = E\psi_2, \quad (3.19)$$

then

$$\hat{H}(a\psi_1 + b\psi_2) = a\hat{H}\psi_1 + b\hat{H}\psi_2 = aE\psi_1 + bE\psi_2 = E(a\psi_1 + b\psi_2). \quad (3.20)$$

We can therefore form two linear combinations of the degenerate eigenstates ψ_n and ψ_{-n} to obtain another description of the energy eigenstates:

$$\begin{aligned} \psi_n + \psi_{-n} &\sim \cos(k_n x), \\ \psi_n - \psi_{-n} &\sim \sin(k_n x). \end{aligned} \quad (3.21)$$

While these are real energy eigenstates, they are not momentum eigenstates. Only our exponentials are simultaneous eigenstates of both \hat{H} and \hat{p} .

The energy eigenstates ψ_n are automatically orthonormal since they are \hat{p} eigenstates with no degeneracies (and as you recall eigenstates of a hermitian operator with different eigenvalues are automatically orthogonal) :

$$\int_0^L \psi_n^*(x) \psi_m(x) dx = \frac{1}{L} \int_0^L e^{\frac{2\pi i (m-n)x}{L}} dx = \delta_{mn}. \quad (3.22)$$

They are also complete: we can then construct a general wavefunction as a superposition that is in fact a Fourier series. For any $\Psi(x, 0)$ that satisfies the periodicity condition, we can write

$$\Psi(x, 0) = \sum_{n \in \mathbb{Z}} a_n \psi_n(x), \quad (3.23)$$

where, as you should check, the coefficients a_n are determined by the integrals

$$a_n = \int_0^L dx \psi_n^*(x) \Psi(x, 0). \quad (3.24)$$

The initial state $\Psi(x, 0)$ is then easily evolved in time:

$$\Psi(x, t) = \sum_{n \in \mathbb{Z}} a_n \psi_n(x) e^{-\frac{i E_n t}{\hbar}}. \quad (3.25)$$

Andrew Turner transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare

<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

LINEAR ALGEBRA: VECTOR SPACES AND OPERATORS

B. Zwiebach

October 21, 2013

Contents

1	Vector spaces and dimensionality	1
2	Linear operators and matrices	5
3	Eigenvalues and eigenvectors	11
4	Inner products	14
5	Orthonormal basis and orthogonal projectors	18
6	Linear functionals and adjoint operators	20
7	Hermitian and Unitary operators	24

1 Vector spaces and dimensionality

In quantum mechanics the state of a physical system is a *vector* in a *complex* vector space. Observables are linear operators, in fact, Hermitian operators acting on this complex vector space. The purpose of this chapter is to learn the basics of vector spaces, the structures that can be built on those spaces, and the operators that act on them.

Complex vector spaces are somewhat different from the more familiar real vector spaces. I would say they have more powerful properties. In order to understand more generally complex vector spaces it is useful to compare them often to their real dimensional friends. We will follow here the discussion of the book *Linear algebra done right*, by Sheldon Axler.

In a vector space one has vectors and numbers. We can add vectors to get vectors and we can multiply vectors by numbers to get vectors. If the numbers we use are real, we have a real vector space. If the numbers we use are complex, we have a complex vector space. More generally, the numbers we use belong to what is called in mathematics a ‘field’ and denoted by the letter \mathbb{F} . We will discuss just two cases, $\mathbb{F} = \mathbb{R}$, meaning that the numbers are real, and $\mathbb{F} = \mathbb{C}$, meaning that the numbers are complex.

The definition of a vector space is the same for \mathbb{F} being \mathbb{R} or \mathbb{C} . A vector space V is a set of vectors with an operation of **addition** (+) that assigns an element $u + v \in V$ to each $u, v \in V$. This means that V is closed under addition. There is also a **scalar multiplication** by elements of \mathbb{F} , with $av \in V$

for any $a \in \mathbb{F}$ and $v \in V$. This means the space V is closed under multiplication by numbers. These operations must satisfy the following additional properties:

1. $u + v = v + u \in V$ for all $u, v \in V$ (addition is commutative).
2. $u + (v + w) = (u + v) + w$ and $(ab)u = a(bu)$ for any $u, v, w \in V$ and $a, b \in \mathbb{F}$ (associativity).
3. There is a vector $0 \in V$ such that $0 + u = u$ for all $u \in V$ (additive identity).
4. For each $v \in V$ there is a $u \in V$ such that $v + u = 0$ (additive inverse).
5. The element $1 \in \mathbb{F}$ satisfies $1v = v$ for all $v \in V$ (multiplicative identity).
6. $a(u + v) = au + av$ and $(a + b)v = av + bv$ for every $u, v \in V$ and $a, b \in \mathbb{F}$ (distributive property).

This definition is very efficient. Several familiar properties follow from it by short proofs (which we will not give, but are not complicated and you may try to produce):

- The additive identity is unique: any vector $0'$ that acts like 0 is actually equal to 0 .
- $0v = 0$, for any $v \in V$, where the first zero is a number and the second one is a vector. This means that the number zero acts as expected when multiplying a vector.
- $a0 = 0$, for any $a \in \mathbb{F}$. Here both zeroes are vectors. This means that the zero vector multiplied by any number is still the zero vector.
- The additive inverse of any vector $v \in V$ is unique. It is denoted by $-v$ and in fact $-v = (-1)v$.

We must emphasize that while the numbers, in \mathbb{F} are sometimes real or complex, we never speak of the vectors themselves as real or complex. A vector multiplied by a complex number is not said to be a complex vector, for example! The vectors in a real vector space are not themselves real, nor are the vectors in a complex vector space complex. We have the following examples of vector spaces:

1. The set of N -component vectors

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}, \quad a_i \in \mathbb{R}, \quad i = 1, 2, \dots, N. \quad (1.1)$$

form a real vector space.

2. The set of $M \times N$ matrices with complex entries

$$\begin{pmatrix} a_{11} & \dots & a_{1N} \\ a_{21} & \dots & a_{2N} \\ \vdots & \vdots & \vdots \\ a_{M1} & \dots & a_{MN} \end{pmatrix}, \quad a_{ij} \in \mathbb{C}, \quad (1.2)$$

is a complex vector space. In here multiplication by a constant multiplies each entry of the matrix by the constant.

3. We can have matrices with complex entries that naturally form a real vector space. The space of two-by-two *hermitian* matrices define a *real* vector space. They do not form a complex vector space since multiplication of a hermitian matrix by a complex number ruins the hermiticity.
4. The set $\mathcal{P}(\mathbb{F})$ of polynomials $p(z)$. Here the variable $z \in \mathbb{F}$ and $p(z) \in \mathbb{F}$. Each polynomial $p(z)$ has coefficients a_0, a_1, \dots, a_n also in \mathbb{F} :

$$p(z) = a_0 + a_1 z + a_2 z^2 + \dots + a_n z^n. \quad (1.3)$$

By definition, the integer n is finite but it can take any nonnegative value. Addition of polynomials works as expected and multiplication by a constant is also the obvious multiplication. The space $\mathcal{P}(\mathbb{F})$ of all polynomials so defined form a vector space over \mathbb{F} .

5. The set \mathbb{F}^∞ of infinite sequences (x_1, x_2, \dots) of elements $x_i \in \mathbb{F}$. Here

$$\begin{aligned} (x_1, x_2, \dots) + (y_1, y_2, \dots) &= (x_1 + y_1, x_2 + y_2, \dots) \\ a(x_1, x_2, \dots) &= (ax_1, ax_2, \dots) \quad a \in \mathbb{F}. \end{aligned} \quad (1.4)$$

This is a vector space over \mathbb{F} .

6. The set of complex functions on an interval $x \in [0, L]$, form a vector space over \mathbb{C} .

To better understand a vector space one can try to figure out its possible subspaces. A **subspace** of a vector space V is a subset of V that is also a vector space. To verify that a subset U of V is a subspace you must check that U contains the vector 0, and that U is closed under addition and scalar multiplication.

Sometimes a vector space V can be described clearly in terms of collection U_1, U_2, \dots, U_m of subspaces of V . We say that the space V is the **direct sum** of the subspaces U_1, U_2, \dots, U_m and we write

$$V = U_1 \oplus U_2 \oplus \dots \oplus U_m \quad (1.5)$$

if any vector in V can be written *uniquely* as the sum $u_1 + u_2 + \dots + u_m$, where $u_i \in U_i$. To check uniqueness one can, alternatively, verify that the only way to write 0 as a sum $u_1 + u_2 + \dots + u_m$ with $u_i \in U_i$ is by taking all u_i 's equal to zero. For the case of two subspaces $V = U \oplus W$, it suffices to prove that any vector can be written as $u + w$ with $u \in U$ and $w \in W$ and that $U \cap W = 0$.

Given a vector space we can produce lists of vectors. A **list** (v_1, v_2, \dots, v_n) of vectors in V contains, by definition, a finite number of vectors. The number of vectors in the list is the length of the list. The **span** of a list of vectors (v_1, v_2, \dots, v_n) in V , denoted as $\text{span}(v_1, v_2, \dots, v_n)$, is the set of all linear combinations of these vectors

$$a_1 v_1 + a_2 v_2 + \dots + a_n v_n, \quad a_i \in \mathbb{F} \quad (1.6)$$

A vector space V is spanned by a list (v_1, v_2, \dots, v_n) if $V = \text{span}(v_1, v_2, \dots, v_n)$.

Now comes a very natural definition: A vector space V is said to be **finite dimensional** if it is spanned by some list of vectors in V . If V is not finite dimensional, it is **infinite dimensional**. In such case, no list of vectors from V can span V .

Let us show that the vector space of all polynomials $p(z)$ considered in Example 4 is an infinite dimensional vector space. Indeed, consider any list of polynomials. In this list there is a polynomial of maximum degree (recall the list is finite). Thus polynomials of higher degree are not in the span of the list. Since no list can span the space, it is infinite dimensional.

For example 1, consider the list of vectors (e_1, e_2, \dots, e_N) with

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots \quad e_N = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \quad (1.7)$$

This list spans the space (the vector displayed is $a_1e_1 + a_2e_2 + \dots + a_Ne_N$). This vector space is finite dimensional.

A list of vectors (v_1, v_2, \dots, v_n) , with $v_i \in V$ is said to be **linearly independent** if the equation

$$a_1v_1 + a_2v_2 + \dots + a_nv_n = 0, \quad (1.8)$$

only has the solution $a_1 = a_2 = \dots = a_n = 0$. One can show that the length of any linearly independent list is shorter or equal to the length of any spanning list. This is reasonable, because spanning lists can be arbitrarily long (adding vectors to a spanning list gives still a spanning list), but a linearly independent list cannot be enlarged beyond a certain point.

Finally, we get to the concept of a basis for a vector space. A **basis** of V is a list of vectors in V that both spans V and it is linearly independent. Mathematicians easily prove that any finite dimensional vector space has a basis. Moreover, all bases of a finite dimensional vector space have the same length. The **dimension** of a finite-dimensional vector space is given by the length of any list of basis vectors. One can also show that for a finite dimensional vector space a list of vectors of length $\dim V$ is a basis if it is linearly independent list or if it is a spanning list.

For example 1 we see that the list (e_1, e_2, \dots, e_N) in (1.7) is not only a spanning list but a linearly independent list (prove it!). Thus the dimensionality of this space is N .

For example 3, recall that the most general hermitian two-by-two matrix takes the form

$$\begin{pmatrix} a_0 + a_3 & a_1 - ia_2 \\ a_1 + ia_2 & a_0 - a_3 \end{pmatrix}, \quad a_0, a_1, a_2, a_3 \in \mathbb{R}. \quad (1.9)$$

Now consider the following list of four ‘vectors’ $(\mathbf{1}, \sigma_1, \sigma_2, \sigma_3)$. All entries in this list are hermitian matrices, so this is a list of vectors in the space. Moreover they span the space since the most general hermitian matrix, as shown above, is simply $a_0\mathbf{1} + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3$. The list is linearly independent

as $a_0\mathbf{1} + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 = 0$ implies that

$$\begin{pmatrix} a_0 + a_3 & a_1 - ia_2 \\ a_1 + ia_2 & a_0 - a_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad (1.10)$$

and you can quickly see that this implies a_0, a_1, a_2 , and a_3 are zero. So the list is a basis and the space in question is a four-dimensional real vector space.

Exercise. Explain why the vector space in example 2 has dimension $M \cdot N$.

It seems pretty obvious that the vector space in example 5 is infinite dimensional, but it actually takes a bit of work to prove it.

2 Linear operators and matrices

A linear map refers in general to a certain kind of function from one vector space V to another vector space W . When the linear map takes the vector space V to itself, we call the linear map a linear operator. We will focus our attention on those operators. Let us then define a linear operator.

A **linear operator** T on a vector space V is a function that takes V to V with the properties:

1. $T(u + v) = Tu + Tv$, for all $u, v \in V$.
2. $T(au) = aTu$, for all $a \in \mathbb{F}$ and $u \in V$.

We call $\mathcal{L}(V)$ the set of all linear operators that act on V . This can be a very interesting set, as we will see below. Let us consider a few examples of linear operators.

1. Let V denote the space of real polynomials $p(x)$ of a real variable x with real coefficients. Here are two linear operators:
 - Let T denote differentiation: $Tp = p'$. This operator is linear because $(p_1 + p_2)' = p'_1 + p'_2$ and $(ap)' = ap'$.
 - Let S denote multiplication by x : $Sp = xp$. S is also a linear operator.
2. In the space \mathbb{F}^∞ of infinite sequences define the left-shift operator L by

$$L(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots). \quad (2.11)$$

We lose the first entry, but that is perfectly consistent with linearity. We also have the right-shift operator R that acts as follows:

$$R(x_1, x_2, \dots) = (0, x_1, x_2, \dots). \quad (2.12)$$

Note that the first entry in the result is zero. It could not be any other number because the zero element (a sequence of all zeroes) should be mapped to itself (by linearity).

3. For any V , the zero map 0 such that $0v = 0$. This map is linear and maps all elements of V to the zero element.
4. For any V , the identity map I for which $Iv = v$ for all $v \in V$. This map leaves all vectors invariant.

Since operators on V can be added and can also be multiplied by numbers, the set $\mathcal{L}(V)$ introduced above is itself a vector space (the vectors being the operators!). Indeed for any two operators $T, S \in \mathcal{L}(V)$ we have the natural definition

$$\begin{aligned} (S + T)v &= Sv + Tv, \\ (aS)v &= a(Sv). \end{aligned} \tag{2.13}$$

The additive identity in the vector space $\mathcal{L}(V)$ is the zero map of example 3.

In this vector space there is a surprising new structure: the vectors (the operators!) can be multiplied. There is a multiplication of linear operators that gives a linear operator. We just let one operator act first and the second later. So given $S, T \in \mathcal{L}(V)$ we define the operator ST as

$$(ST)v \equiv S(Tv) \tag{2.14}$$

You should convince yourself that ST is a linear operator. This product structure in the space of linear operators is associative: $S(TU) = (ST)U$, for S, T, U , linear operators. Moreover it has an identity element: the identity map of example 4. Most crucially this multiplication is, in general, *noncommutative*. We can check this using the two operators T and S of example 1 acting on the polynomial $p = x^n$. Since T differentiates and S multiplies by x we get

$$(TS)x^n = T(Sx^n) = T(x^{n+1}) = (n+1)x^n, \quad \text{while} \quad (ST)x^n = S(Tx^n) = S(nx^{n-1}) = nx^n. \tag{2.15}$$

We can quantify this failure of commutativity by writing the difference

$$(TS - ST)x^n = (n+1)x^n - nx^n = x^n = I x^n \tag{2.16}$$

where we inserted the identity operator at the last step. Since this relation is true for any x^n , it would also hold acting on any polynomial, namely on any element of the vector space. So we write

$$[T, S] = I. \tag{2.17}$$

where we introduced the commutator $[\cdot, \cdot]$ of two operators X, Y , defined as $[X, Y] \equiv XY - YX$.

The most basic features of an operator are captured by two simple concepts: its null space and its range. Given some linear operator T on V it is of interest to consider those elements of V that are mapped to the zero element. The **null space** (or kernel) of $T \in \mathcal{L}(V)$ is the subset of vectors in V that are mapped to zero by T :

$$\text{null } T = \{v \in V; Tv = 0\}. \tag{2.18}$$

Actually $\text{null } T$ is a *subspace* of V (The only nontrivial part of this proof is to show that $T(0) = 0$. This follows from $T(0) = T(0 + 0) = T(0) + T(0)$ and then adding to both sides of this equation the additive inverse to $T(0)$).

A linear operator $T : V \rightarrow V$ is said to be **injective** if $Tu = Tv$, with $u, v \in V$, implies $u = v$. An injective map is called a *one-to-one* map, because not two different elements can be mapped to the same one. In fact, physicist Sean Carroll has suggested that a better name would be *two-to-two* as injectivity really means that two different elements are mapped by T to two different elements! We leave for you as an exercise to prove the following important characterization of injective maps:

Exercise. Show that T is injective if and only if $\text{null } T = \{0\}$.

Given a linear operator T on V it is also of interest to consider the elements of V of the form Tv . The linear operator may not produce by its action all of the elements of V . We define the **range** of T as the image of V under the map T :

$$\text{range } T = \{Tv; v \in V\}. \quad (2.19)$$

Actually $\text{range } T$ is a *subspace* of V (can you prove it?). The linear operator T is said to be **surjective** if $\text{range } T = V$. That is, if the image of V under T is the complete V .

Since both the null space and the range of a linear operator $T : V \rightarrow V$ are subspaces of V , one can assign a dimension to them, and the following theorem is nontrivial:

$$\dim V = \dim(\text{null } T) + \dim(\text{range } T). \quad (2.20)$$

Example. Describe the null space and range of the operator

$$T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (2.21)$$

Let us now consider invertible linear operators. A linear operator $T \in \mathcal{L}(V)$ is **invertible** if there exists another linear operator $S \in \mathcal{L}(V)$ such that ST and TS are identity maps (written as I). The linear operator S is called the **inverse** of T . The inverse is actually unique. Say S and S' are inverses of T . Then we have

$$S = SI = S(TS') = (ST)S' = IS' = S'. \quad (2.22)$$

Note that we required the inverse S to be an inverse acting from the left and acting from the right. This is useful for infinite dimensional vector spaces. For finite-dimensional vector spaces one suffices; one can then show that $ST = I$ if and only if $TS = I$.

It is useful to have a good characterization of invertible linear operators. For a finite-dimensional vector space V the following three statements are equivalent!

$$\text{Finite dimension: } \boxed{T \text{ is invertible}} \longleftrightarrow \boxed{T \text{ is injective}} \longleftrightarrow \boxed{T \text{ is surjective}} \quad (2.23)$$

For infinite dimensional vector spaces injectivity and surjectivity are not equivalent (each can fail independently). In that case invertibility is equivalent to injectivity plus surjectivity:

$$\text{Infinite dimension: } \boxed{T \text{ is invertible}} \longleftrightarrow \boxed{T \text{ is injective and surjective}} \quad (2.24)$$

The left shift operator L is not injective (maps $(x_1, 0, \dots)$ to zero) but it is surjective. The right shift operator is not surjective although it is injective.

Now we consider the **matrix associated to a linear operator T** that acts on a vector space V . This matrix will depend on the basis we choose for V . Let us declare that our basis is the list (v_1, v_2, \dots, v_n) . It is clear that the full knowledge of the action of T on V is encoded in the action of T on the basis vectors, that is on the values $(Tv_1, Tv_2, \dots, Tv_n)$. Since Tv_j is in V , it can be written as a linear combination of basis vectors. We then have

$$\boxed{Tv_j = T_{1j} v_1 + T_{2j} v_2 + \dots + T_{nj} v_n,} \quad (2.25)$$

where we introduced the constants T_{ij} that are known if the operator T is known. As we will see, these are the entries form the matrix representation of the operator T in the chosen basis. The above relation can be written more briefly as

$$\boxed{Tv_j = \sum_{i=1}^n T_{ij} v_i.} \quad (2.26)$$

When we deal with different bases it can be useful to use notation where we replace

$$T_{ij} \rightarrow T_{ij}(\{v\}), \quad (2.27)$$

so that it makes clear that T is being represented using the v basis (v_1, \dots, v_n) .

I want to make clear why (2.25) is reasonable before we show that it makes for a consistent association between operator multiplication and matrix multiplication. The left-hand side, where we have the action of the matrix for T on the j -th basis vector, can be viewed concretely as

$$Tv_j \longleftrightarrow \begin{pmatrix} T_{11} & \cdots & T_{1j} & \cdots & T_{1n} \\ T_{21} & \cdots & T_{2j} & \cdots & T_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{n1} & \cdots & T_{nj} & \cdots & T_{nn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad j\text{-th position} \quad (2.28)$$

where the column vector has zeroes everywhere except on the j -th entry. The product, by the usual rule of matrix multiplication is the column vector

$$\begin{pmatrix} T_{1j} \\ T_{2j} \\ \vdots \\ T_{nj} \end{pmatrix} = T_{1j} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + T_{2j} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \dots + T_{nj} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \longleftrightarrow T_{1j} v_1 + \dots + T_{nj} v_n. \quad (2.29)$$

which we identify with the right-hand side of (2.25). So (2.25) is reasonable.

Exercise. Verify that the matrix representation of the identity operator is a diagonal matrix with an entry of one at each element of the diagonal. This is true for any basis.

Let us now examine the product of two operators and their matrix representation. Consider the operator TS acting on v_j :

$$(TS)v_j = T(Sv_j) = T \sum_p S_{pj} v_p = \sum_p S_{pj} T v_p = \sum_p S_{pj} \sum_i T_{ip} v_i \quad (2.30)$$

so that changing the order of the sums we find

$$(TS)v_j = \sum_i \left(\sum_p T_{ip} S_{pj} \right) v_i. \quad (2.31)$$

Using the identification implicit in (2.26) we see that the object in parenthesis is the i, j matrix element of the matrix that represents TS . Therefore we found

$$(TS)_{ij} = \sum_p T_{ip} S_{pj}, \quad (2.32)$$

which is precisely the right formula for matrix multiplication. In other words, the matrix that represents TS is the product of the matrix that represents T with the matrix that represents S , in that order.

Changing basis

While matrix representations are very useful for concrete visualization, they are basis dependent. It is a good idea to try to figure out if there are quantities that can be calculated using a matrix representation that are, nevertheless, guaranteed to be basis independent. One such quantity is the **trace** of the matrix representation of a linear operator. The trace is the sum of the matrix elements in the diagonal. Remarkably, that sum is the same independent of the basis used. Consider a linear operator T in $\mathcal{L}(V)$ and two sets of basis vectors (v_1, \dots, v_n) and (u_1, \dots, u_n) for V . Using the explicit notation (2.27) for the matrix representation we state this property as

$$\text{tr } T(\{v\}) = \text{tr } T(\{u\}). \quad (2.33)$$

We will establish this result below. On the other hand, if this trace is actually basis independent, there should be a way to define the trace of the linear operator T *without* using its matrix representation. This is actually possible, as we will see. Another basis independent quantity is the determinant of the matrix representation of T .

Let us then consider the effect of a change of basis on the matrix representation of an operator. Consider a vector space V and a change of basis from (v_1, \dots, v_n) to (u_1, \dots, u_n) defined by the linear operator A as follows:

$$A : v_k \rightarrow u_k, \text{ for } k = 1, \dots, n. \quad (2.34)$$

This can also be written as

$$Av_k = u_k \quad (2.35)$$

Since we know how A acts on every element of the basis we know, by linearity how it acts on any vector. The operator A is clearly *invertible* because, letting $B : u_k \rightarrow v_k$ or

$$Bu_k = v_k, \quad (2.36)$$

we have

$$\begin{aligned} BA v_k &= B(Av_k) = Bu_k = v_k \\ AB u_k &= A(Bu_k) = Av_k = u_k, \end{aligned} \quad (2.37)$$

showing that $BA = I$ and $AB = I$. Thus B is the inverse of A . Using the definition of matrix representation, the right-hand sides of the relations $u_k = Av_k$ and $v_k = Bu_k$ can be written so that the equations take the form

$$u_k = A_{jk} v_j, \quad v_k = B_{jk} u_j, \quad (2.38)$$

where we used the convention that repeated indices are summed over. A_{ij} are the elements of the matrix representation of A in the v basis and B_{ij} are the elements of the matrix representation of B in the u basis. Replacing the second relation on the first, and then replacing the first on the second we get

$$\begin{aligned} u_k &= A_{jk} B_{ij} u_i = B_{ij} A_{jk} u_i \\ v_k &= B_{jk} A_{ij} v_i = A_{ij} B_{jk} v_i \end{aligned} \quad (2.39)$$

Since the u 's and v 's are basis vectors we must have

$$B_{ij} A_{jk} = \delta_{ik} \quad \text{and} \quad A_{ij} B_{jk} = \delta_{ik} \quad (2.40)$$

which means that the B matrix is the inverse of the A matrix. We have thus learned that

$$v_k = (A^{-1})_{jk} u_j. \quad (2.41)$$

We can now apply these preparatory results to the matrix representations of the operator T . We have, by definition,

$$Tv_k = T_{ik}(\{v\}) v_i. \quad (2.42)$$

We now want to calculate T on u_k so that we can read the formula for the matrix T on the u basis:

$$Tu_k = T_{ik}(\{u\}) u_i. \quad (2.43)$$

Computing the left-hand side, using the linearity of the operator T , we have

$$Tu_k = T(A_{jk} v_j) = A_{jk} T v_j = A_{jk} T_{pj}(\{v\}) v_p \quad (2.44)$$

and using (2.41) we get

$$Tu_k = A_{jk}T_{pj}(\{v\})(A^{-1})_{ip}u_i = \left((A^{-1})_{ip}T_{pj}(\{v\})A_{jk} \right)u_i = (A^{-1}T(\{v\})A)_{ik}u_i. \quad (2.45)$$

Comparing with (2.43) we get

$$T_{ij}(\{u\}) = (A^{-1}T(\{v\})A)_{ij} \rightarrow \boxed{T(\{u\}) = A^{-1}T(\{v\})A}. \quad (2.46)$$

This is the result we wanted to obtain.

The trace of a matrix T_{ij} is given by T_{ii} , where sum over i is understood. To show that the trace of T is basis independent we write

$$\begin{aligned} \text{tr}(T(\{u\})) &= T_{ii}(\{u\}) = (A^{-1})_{ij}T_{jk}(\{v\})A_{ki} \\ &= A_{ki}(A^{-1})_{ij}T_{jk}(\{v\}) \\ &= \delta_{kj}T_{jk}(\{v\}) = T_{jj}(\{v\}) = \text{tr}(T(\{v\})). \end{aligned} \quad (2.47)$$

For the determinant we recall that $\det(AB) = (\det A)(\det B)$. Therefore $\det(A)\det(A^{-1}) = 1$. From (2.46) we then get

$$\det T(\{u\}) = \det(A^{-1})\det T(\{v\})\det A = \det T(\{v\}). \quad (2.48)$$

Thus the determinant of the matrix that represents a linear operator is independent of the basis used.

3 Eigenvalues and eigenvectors

In quantum mechanics we need to consider eigenvalues and eigenstates of hermitian operators acting on complex vector spaces. These operators are called observables and their eigenvalues represent possible results of a measurement. In order to acquire a better perspective on these matters, we consider the eigenvalue/eigenvector problem in more generality.

One way to understand the action of an operator $T \in \mathcal{L}(V)$ on a vector space V is to understand how it acts on subspaces of V , as those are smaller than V and thus possibly simpler to deal with. Let U denote a subspace of V . In general, the action of T may take elements of U outside U . We have a noteworthy situation if T acting on any element of U gives an element of U . In this case U is said to be **invariant** under T , and T is then a well-defined linear operator on U . A very interesting situation arises if a suitable list of invariant subspaces give the space V as a direct sum.

Of all subspaces, one-dimensional ones are the simplest. Given some vector $u \in V$ one can consider the one-dimensional subspace U spanned by u :

$$U = \{cu : c \in \mathbb{F}\}. \quad (3.49)$$

We can ask if the one-dimensional subspace U is left invariant by the operator T . For this Tu must be equal to a number times u , as this guarantees that $Tu \in U$. Calling the number λ , we write

$$Tu = \lambda u. \quad (3.50)$$

This equation is so ubiquitous that names have been invented to label the objects involved. The number $\lambda \in \mathbb{F}$ is called an **eigenvalue** of the linear operator T if there is a **nonzero** vector $u \in V$ such that the equation above is satisfied. Suppose we find for some specific λ a nonzero vector u satisfying this equation. Then it follows that cu , for any $c \in \mathbb{F}$ also satisfies equation (3.50), so that the solution space of the equation includes the subspace U , which is now said to be an invariant subspace under T . It is convenient to call any vector that satisfies (3.50) for a given λ an **eigenvector** of T corresponding to λ . In doing so we are including the zero vector as a solution and thus as an eigenvector. It can often happen that for a given λ there are several linearly independent eigenvectors. In this case the invariant subspace associated with the eigenvalue λ is higher dimensional. The set of eigenvalues of T is called the **spectrum** of T .

Our equation above is equivalent to

$$(T - \lambda I) u = 0, \quad (3.51)$$

for some nonzero u . It is therefore the case that

λ is an eigenvalue	\longleftrightarrow	$(T - \lambda I)$ not injective.
----------------------------	-----------------------	----------------------------------

(3.52)

Using (2.23) we conclude that λ is an eigenvalue also means that $(T - \lambda I)$ is **not invertible**, and not surjective. We also note that

$$\text{Set of eigenvectors of } T \text{ corresponding to } \lambda = \text{null}(T - \lambda I). \quad (3.53)$$

It should be emphasized that the eigenvalues of T and the invariant subspaces (or eigenvectors associated with fixed eigenvalues) are basis independent objects. Nowhere in our discussion we had to invoke the use of a basis, nor we had to use any matrix representation. Below, we will discuss the familiar calculation of eigenvalues and eigenvectors using a matrix representation of the operator T in some particular basis.

Let us consider some examples. Take a real three-dimensional vector space V (our space to great accuracy!). Consider the rotation operator T that rotates all vectors by a fixed angle small about the z axis. To find eigenvalues and eigenvectors we just think of the invariant subspaces. We must ask which are the vectors for which this rotation doesn't change their direction and effectively just multiplies them by a number? Only the vectors along the z -direction do not change direction upon this rotation. So the vector space spanned by \mathbf{e}_z is the invariant subspace, or the space of eigenvectors. The eigenvectors are associated with the eigenvalue of one, as the vectors are not altered at all by the rotation.

Consider now the case where T is a rotation by ninety degrees on a two-dimensional *real* vector space V . Are there one-dimensional subspaces left invariant by T ? No, **all** vectors are rotated, none remains pointing in the same direction. Thus there are **no eigenvalues**, nor, of course, eigenvectors. If you tried calculating the eigenvalues by the usual recipe, you will find complex numbers. A complex eigenvalue is meaningless in a real vector space.

Although we will not prove the following result, it follows from the facts we have introduced and no extra machinery. It is of interest being completely general and valid for both real and complex vector spaces:

Theorem: Let $T \in \mathcal{L}(V)$ and assume $\lambda_1, \dots, \lambda_n$ are distinct eigenvalues of T and u_1, \dots, u_n are corresponding nonzero eigenvectors. Then (u_1, \dots, u_n) are linearly independent.

Note that we cannot ask if the eigenvectors are orthogonal to each other as we have not yet introduced an inner product on the vector space V . In this theorem there may be more than one linearly independent eigenvector associated with some eigenvalues. In that case any one eigenvector will do. Since an n -dimensional vector space V does not have more than n linearly independent vectors, no linear operator on V can have more than n distinct eigenvalues.

We saw that some linear operators in real vector spaces can fail to have eigenvalues. Complex vector spaces are nicer. In fact, *every linear operator on a finite-dimensional complex vector space has at least one eigenvalue*. This is a fundamental result. It can be proven without using determinants with an elegant argument, but the proof using determinants is quite short.

When λ is an eigenvalue, we have seen that $T - \lambda I$ is not an invertible operator. This also means that using any basis, the matrix representative of $T - \lambda I$ is non-invertible. The condition of non-invertibility of a matrix is identical to the condition that its determinant vanish:

$$\det(T - \lambda \mathbf{1}) = 0. \quad (3.54)$$

This condition, in an N -dimensional vector space looks like

$$\det \begin{pmatrix} T_{11} - \lambda & T_{12} & \dots & T_{1N} \\ T_{21} & T_{22} - \lambda & \dots & T_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ T_{N1} & T_{N2} & \dots & T_{NN} - \lambda \end{pmatrix} = 0. \quad (3.55)$$

The left-hand side is a polynomial $f(\lambda)$ in λ of degree N called the *characteristic polynomial*:

$$f(\lambda) = \det(T - \lambda \mathbf{1}) = (-\lambda)^N + b_{N-1}\lambda^{N-1} + \dots + b_1\lambda + b_0, \quad (3.56)$$

where the b_i are constants. We are interested in the equation $f(\lambda) = 0$, as this determines all possible eigenvalues. If we are working on real vector spaces, the constants b_i are real but there is no guarantee of real roots for $f(\lambda) = 0$. With complex vector spaces, the constants b_i will be complex, but a complex solution for $f(\lambda) = 0$ always exists. Indeed, over the complex numbers we can factor the polynomial $f(\lambda)$ as follows

$$f(\lambda) = (-1)^N (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_N), \quad (3.57)$$

where the notation does not preclude the possibility that some of the λ_i 's may be equal. The λ_i 's are the eigenvalues, since they lead to $f(\lambda) = 0$ for $\lambda = \lambda_i$. If all eigenvalues of T are different the spectrum of T is said to be *non-degenerate*. If an eigenvalue appears k times it is said to be a degenerate eigenvalue with multiplicity k . Even in the most degenerate case we must have at least one eigenvalue. The eigenvectors exist because $(T - \lambda I)$ non-invertible means it is not injective, and therefore there are nonzero vectors that are mapped to zero by this operator.

4 Inner products

We have been able to go a long way without introducing extra structure on the vector spaces. We have considered linear operators, matrix representations, traces, invariant subspaces, eigenvalues and eigenvectors. It is now time to put some additional structure on the vector spaces. In this section we consider a function called an *inner product* that allows us to construct numbers from vectors. A vector space equipped with an inner product is called an inner-product space.

An **inner product** on a vector space V over \mathbb{F} is a machine that takes an *ordered* pair of elements of V , that is, a first vector and a second vector, and yields a number in \mathbb{F} . In order to motivate the definition of an inner product we first discuss the familiar way in which we associate a length to a vector.

The length of a vector, or **norm** of a vector is a real number that is positive or zero, if the vector is the zero vector. In \mathbb{R}^n a vector $a = (a_1, \dots, a_n)$ has norm $|a|$ defined by

$$|a| = \sqrt{a_1^2 + \dots + a_n^2} \quad (4.58)$$

Squaring this one may think of $|a|^2$ as the *dot product* of a with a :

$$|a|^2 = a \cdot a = a_1^2 + \dots + a_n^2 \quad (4.59)$$

Based on this the dot product of any two vectors a and b is defined by

$$a \cdot b = a_1 b_1 + \dots + a_n b_n. \quad (4.60)$$

If we try to generalize this dot product we may require as needed properties the following

1. $a \cdot a \geq 0$, for all vectors a .
2. $a \cdot a = 0$ if and only if $a = 0$.
3. $a \cdot (b_1 + b_2) = a \cdot b_1 + a \cdot b_2$. Additivity in the second entry.
4. $a \cdot (\alpha b) = \alpha a \cdot b$, with α a number.
5. $a \cdot b = b \cdot a$.

Along with these axioms, the length $|a|$ of a vector a is the positive or zero number defined by relation

$$|a|^2 = a \cdot a. \quad (4.61)$$

These axioms are satisfied by the definition (4.60) but do not require it. A new dot product defined by $a \cdot b = c_1 a_1 b_1 + \dots + c_n a_n b_n$, with c_1, \dots, c_n positive constants, would do equally well! So whatever can be proven with these axioms holds true not only for the conventional dot product.

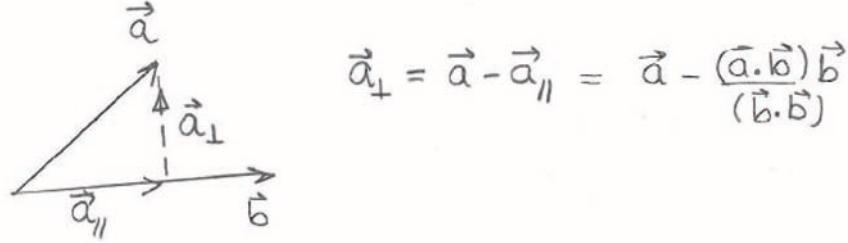
The above axioms guarantee that the Schwarz inequality holds:

$$|a \cdot b| \leq |a| |b|. \quad (4.62)$$

To prove this consider two (nonzero) vectors a and b and then consider the shortest vector joining the tip of a to the line defined by the direction of b (see the figure below). This is the vector a_{\perp} , given by

$$a_{\perp} \equiv a - \frac{a \cdot b}{b \cdot b} b. \quad (4.63)$$

The subscript \perp is there because the vector is perpendicular to b , namely $a_{\perp} \cdot b = 0$, as you can quickly see. To write the above vector we subtracted from a the component of a parallel to b . Note that the vector a_{\perp} is not changed as $b \rightarrow cb$; it does not depend on the overall length of b . Moreover, as it should, the vector a_{\perp} is zero if and only if the vectors a and b are parallel. All this is only motivation, we could have just said “consider the following vector a_{\perp} ”.



Given axiom (1) we have that $a_{\perp} \cdot a_{\perp} \geq 0$ and therefore using (4.63)

$$a_{\perp} \cdot a_{\perp} = a \cdot a - \frac{(a \cdot b)^2}{b \cdot b} \geq 0. \quad (4.64)$$

Since b is not the zero vector we then have

$$(a \cdot b)^2 \leq (a \cdot a)(b \cdot b). \quad (4.65)$$

Taking the square root of this relation we obtain the Schwarz inequality (4.62). The inequality becomes an equality only if $a_{\perp} = 0$ or, as discussed above, when $a = cb$ with c a real constant.

For complex vector spaces some modification is necessary. Recall that the length $|\gamma|$ of a complex number γ is given by $|\gamma| = \sqrt{\gamma^* \gamma}$, where the asterisk superscript denotes complex conjugation. It is

not hard to generalize this a bit. Let $z = (z_1, \dots, z_n)$ be a vector in \mathbb{C}^n . Then the length of the vector $|z|$ is a real number greater than zero given by

$$|z| = \sqrt{z_1^* z_1 + \dots + z_n^* z_n}. \quad (4.66)$$

We must use complex conjugates, denoted by the asterisk superscript, to produce a real number greater than or equal to zero. Squaring this we have

$$|z|^2 = z_1^* z_1 + \dots + z_n^* z_n. \quad (4.67)$$

This suggests that for vectors $z = (z_1, \dots, z_n)$ and $w = (w_1, \dots, w_n)$ an inner product could be given by

$$w_1^* z_1 + \dots + w_n^* z_n, \quad (4.68)$$

and we see that we are not treating the two vectors in an equivalent way. There is the first vector, in this case w whose components are conjugated and a second vector z whose components are not conjugated. If the order of vectors is reversed, we get for the inner product the complex conjugate of the original value. As it was mentioned at the beginning of the section, the inner product requires an ordered pair of vectors. It certainly does for complex vector spaces. Moreover, one can define an inner product in general in a way that applies both to complex and real vector spaces.

An **inner product** on a vector space V over \mathbb{F} is a map from an ordered pair (u, v) of vectors in V to a number $\langle u, v \rangle$ in \mathbb{F} . The axioms for $\langle u, v \rangle$ are inspired by the axioms we listed for the dot product.

1. $\langle v, v \rangle \geq 0$, for all vectors $v \in V$.
2. $\langle v, v \rangle = 0$ if and only if $v = 0$.
3. $\langle u, v_1 + v_2 \rangle = \langle u, v_1 \rangle + \langle u, v_2 \rangle$. Additivity in the second entry.
4. $\langle u, \alpha v \rangle = \alpha \langle u, v \rangle$, with $\alpha \in \mathbb{F}$. Homogeneity in the second entry.
5. $\langle u, v \rangle = \langle v, u \rangle^*$. Conjugate exchange symmetry.

This time the **norm** $|v|$ of a vector $v \in V$ is the positive or zero number defined by relation

$$|v|^2 = \langle v, v \rangle. \quad (4.69)$$

From the axioms above, the only major difference is in number five, where we find that the inner product is not symmetric. We know what complex conjugation is in \mathbb{C} . For the above axioms to apply to vector spaces over \mathbb{R} we just define the obvious: complex conjugation of a real number is a real number. In a real vector space the * conjugation does nothing and the inner product is strictly symmetric in its inputs.

A few comments. One can use (3) with $v_2 = 0$ to show that $\langle u, 0 \rangle = 0$ for all $u \in V$, and thus, by (5) also $\langle 0, u \rangle = 0$. Properties (3) and (4) amount to full linearity in the second entry. It is important to note that additivity holds for the first entry as well:

$$\begin{aligned}\langle u_1 + u_2, v \rangle &= \langle v, u_1 + u_2 \rangle^* \\ &= (\langle v, u_1 \rangle + \langle v, u_2 \rangle)^* \\ &= \langle v, u_1 \rangle^* + \langle v, u_2 \rangle^* \\ &= \langle u_1, v \rangle + \langle u_2, v \rangle.\end{aligned}\tag{4.70}$$

Homogeneity works differently on the first entry, however,

$$\begin{aligned}\langle \alpha u, v \rangle &= \langle v, \alpha u \rangle^* \\ &= (\alpha \langle v, u \rangle)^* \\ &= \alpha^* \langle u, v \rangle.\end{aligned}\tag{4.71}$$

Thus we get **conjugate homogeneity** on the first entry. This is a very important fact. Of course, for a real vector space conjugate homogeneity is the same as just plain homogeneity.

Two vectors $u, v \in V$ are said to be **orthogonal** if $\langle u, v \rangle = 0$. This, of course, means that $\langle v, u \rangle = 0$ as well. The zero vector is orthogonal to all vectors (including itself). Any vector orthogonal to all vectors in the vector space must be equal to zero. Indeed, if $x \in V$ is such that $\langle x, v \rangle = 0$ for all v , pick $v = x$, so that $\langle x, x \rangle = 0$ implies $x = 0$ by axiom 2. This property is sometimes stated as the **non-degeneracy** of the inner product. The “Pythagorean” identity holds for the norm-squared of orthogonal vectors in an inner-product vector space. As you can quickly verify,

$$|u + v|^2 = |u|^2 + |v|^2, \quad \text{for } u, v \in V, \text{ orthogonal vectors.}\tag{4.72}$$

The Schwarz inequality can be proven by an argument fairly analogous to the one we gave above for dot products. The result now reads

Schwarz Inequality: $|\langle u, v \rangle| \leq |u| |v|.$

(4.73)

The inequality is saturated if and only if one vector is a multiple of the other. Note that in the left-hand side $|...|$ denotes the norm of a complex number and on the right-hand side each $|...|$ denotes the norm of a vector. You will prove this identity in a slightly different way in the homework. You will also consider there the *triangle inequality*

$$|u + v| \leq |u| + |v|,\tag{4.74}$$

which is saturated when $u = cv$ for c a real, positive constant. Our definition (4.69) of norm on a vector space V is mathematically sound: a norm is required to satisfy the triangle inequality. Other properties are required: (i) $|v| \geq 0$ for all v , (ii) $|v| = 0$ if and only if $v = 0$, and (iii) $|cv| = |c||v|$ for c some constant. Our norm satisfies all of them.

A complex vector space with an inner product as we have defined is a *Hilbert space* if it is finite dimensional. If the vector space is infinite dimensional, an extra *completeness* requirement must be satisfied for the space to be a Hilbert space: all Cauchy sequences of vectors must converge to vectors in the space. An infinite sequence of vectors v_i , with $i = 1, 2, \dots, \infty$ is a Cauchy sequence if for any $\epsilon > 0$ there is an N such that $|v_n - v_m| < \epsilon$ whenever $n, m > N$.

5 Orthonormal basis and orthogonal projectors

In an inner-product space we can demand that basis vectors have special properties. A list of vectors is said to be **orthonormal** if all vectors have norm one and are pairwise orthogonal. Consider a list (e_1, \dots, e_n) of orthonormal vectors in V . Orthonormality means that

$$\langle e_i, e_j \rangle = \delta_{ij}. \quad (5.75)$$

We also have a simple expression for the norm of $a_1e_1 + \dots + a_ne_n$, with $a_i \in \mathbb{F}$:

$$\begin{aligned} |a_1e_1 + \dots + a_ne_n|^2 &= \langle a_1e_1 + \dots + a_ne_n, a_1e_1 + \dots + a_ne_n \rangle \\ &= \langle a_1e_1, a_1e_1 \rangle + \dots + \langle a_ne_n, a_ne_n \rangle \\ &= |a_1|^2 + \dots + |a_n|^2. \end{aligned} \quad (5.76)$$

This result implies the somewhat nontrivial fact that *the vectors in any orthonormal list are linearly independent*. Indeed if $a_1e_1 + \dots + a_ne_n = 0$ then its norm is zero and so is $|a_1|^2 + \dots + |a_n|^2$. This implies all $a_i = 0$, thus proving the claim.

An **orthonormal basis** of V is a list of orthonormal vectors that is also a basis for V . Let (e_1, \dots, e_n) denote an orthonormal basis. Then any vector v can be written as

$$v = a_1e_1 + \dots + a_ne_n, \quad (5.77)$$

for some constants a_i that can be calculated as follows

$$\langle e_i, v \rangle = \langle e_i, a_1e_1 + \dots + a_ne_n \rangle = a_i, \quad (i \text{ not summed}). \quad (5.78)$$

Therefore any vector v can be written as

$$v = \langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n = \langle e_i, v \rangle e_i. \quad (5.79)$$

To find an orthonormal basis on an inner product space V we just need to start with a basis and then use an algorithm to turn it into an orthogonal basis. In fact, a little more generally:

Gram-Schmidt: Given a list (v_1, \dots, v_n) of linearly independent vectors in V one can construct a list (e_1, \dots, e_n) of orthonormal vectors such that both lists span the same subspace of V .

The Gram-Schmidt algorithm goes as follows. You take e_1 to be v_1 , normalized to have unit norm: $e_1 = v_1/|v_1|$. Then take $v_2 + \alpha e_1$ and fix the constant α so that this vector is orthogonal to e_1 . The

answer is clearly $v_2 - \langle e_1, v_2 \rangle e_1$. This vector, normalized by dividing it by its norm, is set equal to e_2 . In fact we can write the general vector in a recursive fashion. If we know e_1, e_2, \dots, e_{j-1} , we can write e_j as follows:

$$e_j = \frac{v_j - \langle e_1, v_j \rangle e_1 - \dots - \langle e_{j-1}, v_j \rangle e_{j-1}}{|v_j - \langle e_1, v_j \rangle e_1 - \dots - \langle e_{j-1}, v_j \rangle e_{j-1}|} \quad (5.80)$$

It should be clear to you by inspection that this vector is orthogonal to the vectors e_i with $i < j$ and has unit norm. The Gram-Schmidt procedure is quite practical.

With an inner product we can construct interesting subspaces of a vector space V . Consider a subset U of vectors in V (not necessarily a subspace). Then we can define a subspace U^\perp , called the **orthogonal complement** of U as the set of all vectors orthogonal to the vectors in U :

$$U^\perp = \{v \in V \mid \langle v, u \rangle = 0, \text{ for all } u \in U\}. \quad (5.81)$$

This is clearly a subspace of V . When U is a subspace, then U and U^\perp actually give a direct sum decomposition of the full space:

Theorem: If U is a subspace of V , then $V = U \oplus U^\perp$.

Proof: This is a fundamental result and is not hard to prove. Let (e_1, \dots, e_n) be an orthonormal basis for U . We can clearly write any vector v in V as

$$v = (\langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n) + (v - \langle e_1, v \rangle e_1 - \dots - \langle e_n, v \rangle e_n). \quad (5.82)$$

On the right-hand side the first vector in parenthesis is clearly in U as it is written as a linear combination of U basis vectors. The second vector is clearly in U^\perp as one can see that it is orthogonal to any vector in U . To complete the proof one must show that there is no vector except the zero vector in the intersection $U \cap U^\perp$ (recall the comments below (1.5)). Let $v \in U \cap U^\perp$. Then v is in U and in U^\perp so it should satisfy $\langle v, v \rangle = 0$. But then $v = 0$, completing the proof.

Given this decomposition any vector $v \in V$ can be written uniquely as $v = u + w$ where $u \in U$ and $w \in U^\perp$. One can define a linear operator P_U , called the **orthogonal projection** of V onto U , that acting on v above gives the vector u . It is clear from this definition that: (i) the range of P_U is U , (ii) the null space of P_U is U^\perp , (iii) that P_U is not invertible and, (iv) acting on U , the operator P_U is the identity operator. The formula for the vector u can be read from (5.82)

$$P_U v = \langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n. \quad (5.83)$$

It is a straightforward but a good exercise to verify that this formula is consistent with the fact that acting on U , the operator P_U is the identity operator. Thus if we act twice in succession with P_U on a vector, the second action has no effect as it is already acting on a vector in U . It follows from this that

$$P_U P_U = I P_U = P_U \rightarrow \boxed{P_U^2 = P_U}. \quad (5.84)$$

The eigenvalues and eigenvectors of P_U are easy to describe. Since all vectors in U are left invariant by the action of P_U , an orthonormal basis of U provides a set of orthonormal eigenvectors of P all with

eigenvalue one. If we choose on U^\perp an orthonormal basis, that basis provides orthonormal eigenvectors of P all with eigenvalue zero.

In fact equation (5.84) implies that the eigenvalues of P_U can only be one or zero. The eigenvalues of an operator satisfy whatever equation the operator satisfies (as shown by letting the equation act on a presumed eigenvector) thus $\lambda^2 = \lambda$ is needed, and this gives $\lambda(\lambda - 1) = 0$, and $\lambda = 0, 1$, as the only possibilities.

Consider a vector space $V = U \oplus U^\perp$ that is $(n + k)$ -dimensional, where U is n -dimensional and U^\perp is k -dimensional. Let (e_1, \dots, e_n) be an orthonormal basis for U and (f_1, \dots, f_k) an orthonormal basis for U^\perp . We then see that the list of vectors (g_1, \dots, g_{n+k}) defined by

$$(g_1, \dots, g_{n+k}) = (e_1, \dots, e_n, f_1, \dots, f_k) \text{ is an orthonormal basis for } V. \quad (5.85)$$

Exercise: Use $P_U e_i = e_i$, for $i = 1, \dots, n$ and $P_U f_i = 0$, for $i = 1, \dots, k$, to show that in the above basis the projector operator is represented by the diagonal matrix:

$$P_U = \text{diag}\left(\underbrace{1, \dots, 1}_{n \text{ entries}}, \underbrace{0, \dots, 0}_{k \text{ entries}}\right). \quad (5.86)$$

We see that, as expected from its non-invertibility, $\det(P_U) = 0$. But more interestingly we see that the trace of the matrix P_U is n . Therefore

$$\text{tr } P_U = \dim U. \quad (5.87)$$

The dimension of U is the **rank** of the projector P_U . Rank one projectors are the most common projectors. They project to one-dimensional subspaces of the vector space.

Projection operators are useful in quantum mechanics, where observables are described by operators. The effect of measuring an observable on a physical state vector is to turn this original vector instantaneously into another vector. This resulting vector is the orthogonal projection of the original vector down to some eigenspace of the operator associated with the observable.

6 Linear functionals and adjoint operators

When we consider a linear operator T on a vector space V that has an inner product, we can construct a related linear operator T^\dagger on V called the **adjoint** of T . This is a very useful operator and is typically different from T . When the adjoint T^\dagger happens to be equal to T , the operator is said to be *Hermitian*. To understand adjoints, we first need to develop the concept of a linear functional.

A **linear functional** ϕ on the vector space V is a linear map from V to the numbers \mathbb{F} : for $v \in V$, $\phi(v) \in \mathbb{F}$. A linear functional has the following two properties:

1. $\phi(v_1 + v_2) = \phi(v_1) + \phi(v_2)$, with $v_1, v_2 \in V$.
2. $\phi(av) = a\phi(v)$ for $v \in V$ and $a \in \mathbb{F}$.

As an example, consider the three-dimensional real vector space \mathbb{R}^3 with inner product equal to the familiar dot product. Writing a vector v as the triplet $v = (v_1, v_2, v_3)$, we take

$$\phi(v) = 3v_1 + 2v_2 - 4v_3. \quad (6.1)$$

Linearity is clear as the right-hand side features the components v_1, v_2, v_3 appearing linearly. We can use a vector $u = (3, 2, -4)$ to write the linear functional as an inner product. Indeed, one can readily see that

$$\phi(v) = \langle u, v \rangle. \quad (6.2)$$

This is no accident, in fact. We can prove that any linear functional $\phi(v)$ admits such representation with some suitable choice of vector u .

Theorem: Let ϕ be a linear functional on V . There is a unique vector $u \in V$ such that $\phi(v) = \langle u, v \rangle$ for all $v \in V$.

Proof: Consider an orthonormal basis, (e_1, \dots, e_n) and write the vector v as

$$v = \langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n. \quad (6.3)$$

When ϕ acts on v we find, first by linearity and then by conjugate homogeneity

$$\begin{aligned} \phi(v) &= \phi(\langle e_1, v \rangle e_1 + \dots + \langle e_n, v \rangle e_n) \\ &= \langle e_1, v \rangle \phi(e_1) + \dots + \langle e_n, v \rangle \phi(e_n) \\ &= \langle \phi(e_1)^* e_1, v \rangle + \dots + \langle \phi(e_n)^* e_n, v \rangle \\ &= \langle \phi(e_1)^* e_1 + \dots + \phi(e_n)^* e_n, v \rangle. \end{aligned} \quad (6.4)$$

We have thus shown that, as claimed

$$\phi(v) = \langle u, v \rangle \quad \text{with} \quad u = \phi(e_1)^* e_1 + \dots + \phi(e_n)^* e_n. \quad (6.5)$$

Next, we prove that this u is unique. If there exists another vector, u' , that also gives the correct result for all v , then $\langle u', v \rangle = \langle u, v \rangle$, which implies $\langle u - u', v \rangle = 0$ for all v . Taking $v = u' - u$, we see that this shows $u' - u = 0$ or $u' = u$, proving uniqueness.¹

We can modify a bit the notation when needed, to write

$$\phi_u(v) \equiv \langle u, v \rangle, \quad (6.6)$$

where the left-hand side makes it clear that this is a functional acting on v that depends on u .

We can now address the construction of the adjoint. Consider: $\phi(v) = \langle u, T v \rangle$, which is clearly a linear functional, whatever the operator T is. Since any linear functional can be written as $\langle w, v \rangle$, with some suitable vector w , we write

$$\langle u, T v \rangle = \langle w, v \rangle, \quad (6.7)$$

¹This theorem holds for infinite dimensional Hilbert spaces, for *continuous* linear functionals.

Of course, the vector w must depend on the vector u that appears on the left-hand side. Moreover, it must have something to do with the operator T , who does not appear anymore on the right-hand side. So we must look for some good notation here. We can think of w as a function of the vector u and thus write $w = T^\dagger u$ where T^\dagger denotes a map (not obviously linear) from V to V . So, we think of $T^\dagger u$ as the vector obtained by acting with some function T^\dagger on u . The above equation is written as

$$\langle u, Tv \rangle = \langle T^\dagger u, v \rangle, \quad (6.8)$$

Our next step is to show that, in fact, T^\dagger is a linear operator on V . The operator T^\dagger is called the **adjoint** of T . Consider

$$\langle u_1 + u_2, Tv \rangle = \langle T^\dagger(u_1 + u_2), v \rangle, \quad (6.9)$$

and work on the left-hand side to get

$$\begin{aligned} \langle u_1 + u_2, Tv \rangle &= \langle u_1, Tv \rangle + \langle u_2, Tv \rangle \\ &= \langle T^\dagger u_1, v \rangle + \langle T^\dagger u_2, v \rangle \\ &= \langle T^\dagger u_1 + T^\dagger u_2, v \rangle. \end{aligned} \quad (6.10)$$

Comparing the right-hand sides of the last two equations we get the desired:

$$T^\dagger(u_1 + u_2) = T^\dagger u_1 + T^\dagger u_2. \quad (6.11)$$

Having established linearity now we establish homogeneity. Consider

$$\langle au, Tv \rangle = \langle T^\dagger(au), v \rangle. \quad (6.12)$$

The left hand side is

$$\langle au, Tv \rangle = a^* \langle u, Tv \rangle = a^* \langle T^\dagger u, v \rangle = \langle aT^\dagger u, v \rangle. \quad (6.13)$$

This time we conclude that

$$T^\dagger(au) = aT^\dagger u. \quad (6.14)$$

This concludes the proof that T^\dagger , so defined is a linear operator on V .

A couple of important properties are readily proven:

Claim: $(ST)^\dagger = T^\dagger S^\dagger$. We can show this as follows: $\langle u, STv \rangle = \langle S^\dagger u, Tv \rangle = \langle T^\dagger S^\dagger u, v \rangle$.

Claim: The adjoint of the adjoint is the original operator: $(S^\dagger)^\dagger = S$. We can show this as follows: $\langle u, S^\dagger v \rangle = \langle (S^\dagger)^\dagger u, v \rangle$. Now, additionally $\langle u, S^\dagger v \rangle = \langle S^\dagger v, u \rangle^* = \langle v, Su \rangle^* = \langle Su, v \rangle$. Comparing with the first result, we have shown that $(S^\dagger)^\dagger u = Su$, for any u , which proves the claim

Example: Let $v = (v_1, v_2, v_3)$, with $v_i \in \mathbb{C}$ denote a vector in the three-dimensional complex vector space, \mathbb{C}^3 . Define a linear operator T that acts on v as follows:

$$T(v_1, v_2, v_3) = (0v_1 + 2v_2 + iv_3, v_1 - iv_2 + 0v_3, 3iv_1 + v_2 + 7v_3). \quad (6.15)$$

Calculate the action of T^\dagger on a vector. Give the matrix representations of T and T^\dagger using the orthonormal basis $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$. Assume the inner product is the standard one on \mathbb{C}^3 .

Solution: We introduce a vector $u = (u_1, u_2, u_3)$ and will use the basic identity $\langle u, Tv \rangle = \langle T^\dagger u, v \rangle$. The left-hand side of the identity gives:

$$\langle u, Tv \rangle = u_1^*(2v_2 + iv_3) + u_2^*(v_1 - iv_2) + u_3^*(3iv_1 + v_2 + 7v_3). \quad (6.16)$$

This is now rewritten by factoring the various v_i 's

$$\langle u, Tv \rangle = (u_2^* + 3iu_3)v_1 + (2u_1^* - iu_2^* + u_3^*)v_2 + (iu_1^* + 7u_3^*)v_3. \quad (6.17)$$

Identifying the right-hand side with $\langle T^\dagger u, v \rangle$ we now deduce that

$$T^\dagger(u_1, u_2, u_3) = (u_2 - 3iu_3, 2u_1 + iu_2 + u_3, -iu_1 + 7u_3). \quad (6.18)$$

This gives the action of T^\dagger . To find the matrix representation we begin with T . Using basis vectors, we have from (6.15)

$$Te_1 = T(1, 0, 0) = (0, 1, 3i) = e_2 + 3ie_3 = T_{11}e_1 + T_{21}e_2 + T_{31}e_3, \quad (6.19)$$

and deduce that $T_{11} = 0$, $T_{21} = 1$, $T_{31} = 3i$. This can be repeated, and the rule becomes clear quickly: the coefficients of v_i read left to right fit into the i -th column of the matrix. Thus, we have

$$T = \begin{pmatrix} 0 & 2 & i \\ 1 & -i & 0 \\ 3i & 1 & 7 \end{pmatrix} \quad \text{and} \quad T^\dagger = \begin{pmatrix} 0 & 1 & -3i \\ 2 & i & 1 \\ -i & 0 & 7 \end{pmatrix}. \quad (6.20)$$

These matrices are related: one is the transpose and complex conjugate of the other! This is not an accident.

Let us reframe this using matrix notation. Let $u = e_i$ and $v = e_j$ where e_i and e_j are orthonormal basis vectors. Then the definition $\langle u, Tv \rangle = \langle T^\dagger u, v \rangle$ can be written as

$$\begin{aligned} \langle T^\dagger e_i, e_j \rangle &= \langle e_i, Te_j \rangle \\ \langle T_{ki}^\dagger e_k, e_j \rangle &= \langle e_i, T_{kj} e_k \rangle \\ (T_{ki}^\dagger)^* \delta_{kj} &= T_{jk} \delta_{ik} \\ (T^\dagger)_{ji}^* &= T_{ij} \end{aligned} \quad (6.21)$$

Relabeling i and j and taking the complex conjugate we find the familiar relation between a matrix and its adjoint:

$$(T^\dagger)_{ij} = (T_{ji})^*. \quad (6.22)$$

The adjoint matrix is the transpose and complex conjugate matrix only if we use an orthonormal basis. If we did not, in the equation above the use of $\langle e_i, e_j \rangle = \delta_{ij}$ would be replaced by $\langle e_i, e_j \rangle = g_{ij}$, where g_{ij} is some constant matrix that would appear in the rule for the construction of the adjoint matrix.

7 Hermitian and Unitary operators

Before we begin looking at special kinds of operators let us consider a very surprising fact about operators on complex vector spaces, as opposed to operators on real vector spaces.

Suppose we have an operator T that is such that for any vector $v \in V$ the following inner product vanishes

$$\langle v, T v \rangle = 0 \quad \text{for all } v \in V. \quad (7.23)$$

What can we say about the operator T ? The condition states that T is an operator that starting from a vector gives a vector orthogonal to the original one. In a two-dimensional real vector space, this is simply the operator that rotates any vector by ninety degrees! It is quite surprising and important that for *complex* vector spaces the result is very strong: any such operator T necessarily vanishes. This is a theorem:

Theorem: Let T be a linear operator in a **complex vector space** V :

$\text{If } \langle v, T v \rangle = 0 \text{ for all } v \in V, \text{ then } T = 0.$

(7.24)

Proof: Any proof must be such that it fails to work for real vector space. Note that the result follows if we could prove that $\langle u, T v \rangle = 0$, for all $u, v \in V$. Indeed, if this holds, then take $u = T v$, then $\langle T v, T v \rangle = 0$ for all v implies that $T v = 0$ for all v and therefore $T = 0$.

We will thus try to show that $\langle u, T v \rangle = 0$ for all $u, v \in V$. All we know is that objects of the form $\langle \#, T \# \rangle$ vanish, whatever $\#$ is. So we must aim to form linear combinations of such terms in order to reproduce $\langle u, T v \rangle$. We begin by trying the following

$$\langle u + v, T(u + v) \rangle - \langle u - v, T(u - v) \rangle = 2\langle u, T v \rangle + 2\langle v, T u \rangle. \quad (7.25)$$

We see that the “diagonal” term vanished, but instead of getting just $\langle u, T v \rangle$ we also got $\langle v, T u \rangle$. Here is where complex numbers help, we can get the same two terms but with opposite signs by trying,

$$\langle u + iv, T(u + iv) \rangle - \langle u - iv, T(u - iv) \rangle = 2i\langle u, T v \rangle - 2i\langle v, T u \rangle. \quad (7.26)$$

It follows from the last two relations that

$$\langle u, T v \rangle = \frac{1}{4} \left(\langle u + v, T(u + v) \rangle - \langle u - v, T(u - v) \rangle + \frac{1}{i} \langle u + iv, T(u + iv) \rangle - \frac{1}{i} \langle u - iv, T(u - iv) \rangle \right). \quad (7.27)$$

The condition $\langle v, T v \rangle = 0$ for all v , implies that each term of the above right-hand side vanishes, thus showing that $\langle u, T v \rangle = 0$ for all $u, v \in V$. As explained above this proves the result.

An operator T is said to be **Hermitian** if $T^\dagger = T$. Hermitian operators are pervasive in quantum mechanics. The above theorem in fact helps us discover Hermitian operators. It is familiar that the expectation value of a Hermitian operator, on any state, is real. It is also true, however, that any operator whose expectation value is real for all states must be Hermitian:

$$T = T^\dagger \text{ if and only if } \langle v, Tv \rangle \in \mathbb{R} \text{ for all } v. \quad (7.28)$$

To prove this first go from left to right. If $T = T^\dagger$

$$\langle v, Tv \rangle = \langle T^\dagger v, v \rangle = \langle Tv, v \rangle = \langle v, Tv \rangle^*, \quad (7.29)$$

showing that $\langle v, Tv \rangle$ is real. To go from right to left first note that the reality condition means that

$$\langle v, Tv \rangle = \langle Tv, v \rangle = \langle v, T^\dagger v \rangle, \quad (7.30)$$

where the last equality follows because $(T^\dagger)^\dagger = T$. Now the leftmost and rightmost terms can be combined to give $\langle v, (T - T^\dagger)v \rangle = 0$, which holding for all v implies, by the theorem, that $T = T^\dagger$.

We can prove two additional results of Hermitian operators rather easily. We have discussed earlier the fact that on a complex vector space any linear operator has at least one eigenvalue. Here we learn that the eigenvalues of a hermitian operator are real numbers. Moreover, while we have noted that eigenvectors corresponding to different eigenvalues are linearly independent, for Hermitian operators they are guaranteed to be orthogonal. Thus we have the following theorems

Theorem 1: The eigenvalues of Hermitian operators are real.

Theorem 2: Different eigenvalues of a Hermitian operator correspond to orthogonal eigenfunctions.

Proof 1: Let v be a nonzero eigenvector of the Hermitian operator T with eigenvalue λ : $Tv = \lambda v$. Taking the inner product with v we have that

$$\langle v, Tv \rangle = \langle v, \lambda v \rangle = \lambda \langle v, v \rangle. \quad (7.31)$$

Since T is hermitian, we can also evaluate $\langle v, Tv \rangle$ as follows

$$\langle v, Tv \rangle = \langle Tv, v \rangle = \langle \lambda v, v \rangle = \lambda^* \langle v, v \rangle. \quad (7.32)$$

The above equations give $(\lambda - \lambda^*) \langle v, v \rangle = 0$ and since v is not the zero vector, we conclude that $\lambda^* = \lambda$, showing the reality of λ .

Proof 2: Let v_1 and v_2 be eigenvectors of the operator T :

$$Tv_1 = \lambda_1 v_1, \quad Tv_2 = \lambda_2 v_2, \quad (7.33)$$

with λ_1 and λ_2 real (previous theorem) and different from each other. Consider the inner product $\langle v_2, Tv_1 \rangle$ and evaluate it in two different ways. First

$$\langle v_2, Tv_1 \rangle = \langle v_2, \lambda_1 v_1 \rangle = \lambda_1 \langle v_2, v_1 \rangle, \quad (7.34)$$

and second, using hermiticity of T ,

$$\langle v_2, T v_1 \rangle = \langle T v_2, v_1 \rangle = \langle \lambda_2 v_2, v_1 \rangle = \lambda_2 \langle v_2, v_1 \rangle. \quad (7.35)$$

From these two evaluations we conclude that

$$(\lambda_1 - \lambda_2) \langle v_1, v_2 \rangle = 0 \quad (7.36)$$

and the assumption $\lambda_1 \neq \lambda_2$, leads to $\langle v_1, v_2 \rangle = 0$, showing the orthogonality of the eigenvectors.

Let us now consider another important class of linear operators on a complex vector space, the so-called unitary operators. An operator $U \in \mathcal{L}(V)$ in a complex vector space V is said to be a **unitary operator** if it is surjective and does not change the magnitude of the vector it acts upon:

$$|Uu| = |u|, \text{ for all } u \in V. \quad (7.37)$$

We tailored the definition to be useful even for infinite dimensional spaces. Note that U can only kill vectors of zero length, and since the only such vector is the zero vector, $\text{null } U = 0$, and U is injective. Since U is also assumed to be surjective, a **unitary operator U is always invertible**.

A simple example of a unitary operator is the operator λI with λ a complex number of unit-norm: $|\lambda| = 1$. Indeed $|\lambda Iu| = |\lambda u| = |\lambda||u| = |u|$ for all u . Moreover, the operator is clearly surjective.

For another useful characterization of unitary operators we begin by squaring (7.37)

$$\langle Uu, Uu \rangle = \langle u, u \rangle \quad (7.38)$$

By the definition of adjoint

$$\langle u, U^\dagger U u \rangle = \langle u, u \rangle \rightarrow \langle u, (U^\dagger U - I)u \rangle = 0 \text{ for all } u. \quad (7.39)$$

So by our theorem $U^\dagger U = I$, and since U is invertible this means U^\dagger is the inverse of U and we also have $UU^\dagger = I$:

$$U^\dagger U = UU^\dagger = I.$$

(7.40)

Unitary operators *preserve inner products* in the following sense

$$\langle Uu, Uv \rangle = \langle u, v \rangle. \quad (7.41)$$

This follows immediately by moving the second U to act on the first input and using $U^\dagger U = I$.

Assume the vector space V is finite dimensional and has an orthonormal basis (e_1, \dots, e_n) . Consider the new set of vectors (f_1, \dots, f_n) where the f 's are obtained from the e 's by the action of a unitary operator U :

$$f_i = U e_i. \quad (7.42)$$

This also means that $e_i = U^\dagger f_i$. We readily see that the f 's are also a basis, because they are linearly independent: Acting on $a_1 f_1 + \dots + a_n f_n = 0$ with U^\dagger we find $a_1 e_1 + \dots + a_n e_n = 0$, and thus $a_i = 0$. We now see that the new basis is also orthonormal:

$$\langle f_i, f_j \rangle = \langle U e_i, U e_j \rangle = \langle e_i, e_j \rangle = \delta_{ij}. \quad (7.43)$$

The matrix elements of U in the e -basis are

$$U_{ki} = \langle e_k, U e_i \rangle. \quad (7.44)$$

Let us compute the matrix elements U'_{ki} of U in the f -basis

$$U'_{ki} = \langle f_k, U f_i \rangle = \langle U e_k, U f_i \rangle = \langle e_k, f_i \rangle = \langle e_k, U e_i \rangle = U_{ki} \quad (7.45)$$

The matrix elements are the same! Can you find an explanation for this result?

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 4

B. Zwiebach
February 18, 2016

Contents

1 de Broglie wavelength and Galilean transformations	1
2 Phase and Group Velocities	4
3 Choosing the wavefunction for a free particle	6

1 de Broglie wavelength and Galilean transformations

We have seen that to any free particle with momentum \mathbf{p} , we can associate a plane wave, or a “matter wave”, with de Broglie wavelength $\lambda = h/p$, with $p = |\mathbf{p}|$. The question is, *waves of what?* Well, this wave is eventually recognized as an example of what one calls the *wavefunction*. The wavefunction, as we will see is governed by the Schrödinger equation. As we have hinted, the wavefunction gives us information about probabilities, and we will develop this idea in detail.

Does the wave have directional or polarization properties like electric and magnetic fields in an electromagnetic wave? Yes, there is an analog of this, although we will not delve into it now. The analog of polarization corresponds to spin! The effects of spin are negligible in many cases (small velocities, no magnetic fields, for example) and for this reason, we just use a scalar wave, a complex number

$$\Psi(\mathbf{x}, t) \in \mathbb{C} \quad (1.1)$$

that depends on space and time. A couple of obvious questions come to mind. Is the wavefunction measurable? What kind of object is it? What does it describe? In trying to get intuition about this, let's consider how different observers perceive the de Broglie wavelength of a particle, which should help us understand what kind of waves we are talking about. Recall that

$$p = \frac{h}{\lambda} = \frac{h}{2\pi} \frac{2\pi}{\lambda} = \hbar k, \quad (1.2)$$

where k is the wavenumber. How would this wave behave under a change of frame?

We therefore consider two frames S and S' with the x and x' axes aligned and with S' moving to the right along the $+x$ direction of S with constant velocity v . At time equal zero, the origins of the two reference frames coincide.

The time and spatial coordinates of the two frames are related by a *Galilean transformation*, which states that

$$x' = x - vt, \quad t' = t. \quad (1.3)$$

Indeed time runs at the same speed in all Galilean frames and the relation between x and x' is manifest from the arrangement shown in Fig. 1.

Now assume both observers focus on a particle of mass m moving with nonrelativistic speed. Call the speed and momentum in the S frame \tilde{v} and $p = m\tilde{v}$, respectively. It follows by differentiation with

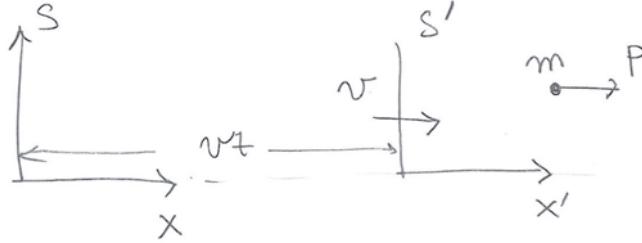


Figure 1: The S' frame moves at speed v along the x -direction of the S frame. A particle of mass m moves with speed \tilde{v} , and thus momentum $p = m\tilde{v}$, in the S frame.

respect to $t = t'$ of the first equation in (1.3) that

$$\frac{dx'}{dt'} = \frac{dx}{dt} - v, \quad (1.4)$$

which means that the particle velocity \tilde{v}' in the S' frame is given by

$$\tilde{v}' = \tilde{v} - v. \quad (1.5)$$

Multiplying by the mass m we find the relation between the momenta in the two frames

$$p' = p - mv. \quad (1.6)$$

The momentum p' in the S' frame can be appreciably different from the momentum p in the S frame. Thus the observers in S' and in S will obtain rather different de Broglie wavelengths λ' and λ ! Indeed,

$$\lambda' = \frac{h}{p'} = \frac{h}{p - mv} \neq \lambda, \quad (1.7)$$

This is very strange! As we review now, for ordinary waves that propagate in the rest frame of a medium (like sound waves or water waves) Galilean observers will find frequency changes but no change in wavelength. This is intuitively clear: to find the wavelength one need only take a picture of the wave at some given time, and both observers looking at the picture will agree on the value of the wavelength. On the other hand to measure frequency, each observers must wait some time to see a full period of the wave go through them. This will take different time for the different observers.

Let us demonstrate these claims quantitatively. We begin with the statement that the phase $\phi = kx - \omega t$ of such a wave is a Galilean invariant. The wave itself may be $\cos \phi$ or $\sin \phi$ or some combination, but the fact is that the physical value of the wave at any point and time must be agreed by the two observers. The wave is an observable. Since all the features of the wave (peaks, zeroes, etc, etc) are controlled by the phase, the two observers must agree on the value of the phase.

In the S frame the phase can be written as follows

$$\phi = kx - \omega t = k(x - \frac{\omega}{k}t) = \frac{2\pi}{\lambda}(x - Vt) = \frac{2\pi x}{\lambda} - \frac{2\pi V}{\lambda}t, \quad (1.8)$$

where $V = \frac{\omega}{k}$ is the wave velocity. Note that the wavelength is read from the coefficient of x and ω is minus the coefficient of t . The two observers should agree on the value of ϕ . That is, we should have

$$\phi'(x', t') = \phi(x, t) \quad (1.9)$$

where the coordinates and times are related by a Galilean transformation. Therefore

$$\phi'(x', t') = \frac{2\pi}{\lambda}(x - Vt) = \frac{2\pi}{\lambda}(x' + vt' - Vt') = \frac{2\pi}{\lambda}x' - \frac{2\pi(V-v)}{\lambda}t'. \quad (1.10)$$

Since the right-hand side is expressed in terms of primed variables, we can read λ' from the coefficient of x' and ω' as minus the coefficient of t' :

$$\lambda' = \lambda \quad (1.11)$$

$$\omega' = \frac{2\pi}{\lambda}(V - v) = \frac{2\pi V}{\lambda} \left(1 - \frac{v}{V}\right) = \omega \left(1 - \frac{v}{V}\right). \quad (1.12)$$

This confirms that, as we claimed, for a physical wave propagating in a medium, the wavelength is a Galilean invariant and the frequency transforms.

So what does it mean that the wavelength of matter waves change under a Galilean transformation? It means that the Ψ waves are not directly measurable! Their value does not correspond to a measurable quantity for which all Galilean observers must agree. Thus, the wavefunction need not be invariant under Galilean transformations:

$$\Psi(x, t) \neq \Psi'(x', t'), \quad (1.13)$$

where (x, t) and (x', t') are related by Galilean transformations and thus represent the same point and time. You will figure out in Homework the correct relation between $\Psi(x, t)$ and $\Psi'(x', t')$.

What is the frequency ω of the de Broglie wave for a particle with momentum p ? We had

$$p = \hbar k \quad (1.14)$$

which fixes the wavelength in terms of the momentum. The frequency ω of the wave is determined by the relation

$$E = \hbar\omega, \quad (1.15)$$

which was also postulated by de Broglie and fixes ω in terms of the energy E of the particle. Note that for our focus on non-relativistic particles the energy E is determined by the momentum through the relation

$$E = \frac{p^2}{2m}. \quad (1.16)$$

We can give three pieces of evidence that (1.15) is a reasonable relation.

1. If we superpose matter waves to form a wave-packet that represents the particle, the packet will move with the so called group velocity v_g , which in fact coincides with the velocity of the particle. The group velocity is found by differentiation of ω with respect to k , as we will review soon:

$$v_g = \frac{d\omega}{dk} = \frac{dE}{dp} = \frac{d}{dp} \left(\frac{p^2}{2m} \right) = \frac{p}{m} = v. \quad (1.17)$$

2. The relation is also suggested by special relativity. The energy and the momentum components of a particle form a four-vector:

$$\left(\frac{E}{c}, p \right) \quad (1.18)$$

Similarly, for waves whose phases are relativistic invariant we have another four-vector

$$\left(\frac{\omega}{c}, k \right) \quad (1.19)$$

Setting two four-vectors equal to each other is a consistent choice: it would be valid in all Lorentz frames. As you can see, both de Broglie relations follow from

$$\left(\frac{E}{c}, p \right) = \hbar \left(\frac{\omega}{c}, k \right). \quad (1.20)$$

3. For photons, (1.15) is consistent with Einstein's quanta of energy, because $E = h\nu = \hbar\omega$.

In summary we have

$$p = \hbar k, \quad E = \hbar\omega. \quad (1.21)$$

These are called the *de Broglie relations*, and they are valid for all particles.

2 Phase and Group Velocities

To understand group velocity we form wave packets and investigate how fast they move. For this we will simply assume that $\omega(k)$ is some arbitrary function of k . Consider a superposition of plane waves $e^{i(kx-\omega(k)t)}$ given by

$$\psi(x, t) = \int dk \Phi(k) e^{i(kx-\omega(k)t)}. \quad (2.22)$$

We assume that the function $\Phi(k)$ is peaked around some wavenumber $k = k_0$, as shown in Fig. 2.

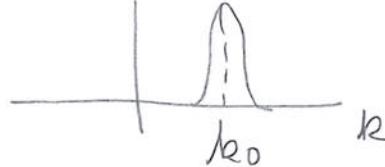


Figure 2: The function $\Phi(k)$ is assumed to peak around $k = k_0$.

In order to motivate the following discussion consider the case when $\Phi(k)$ not only peaks around k_0 but it also is *real* (we will drop this assumption later). In this case the phase φ of the integrand comes only from the exponential:

$$\varphi(k) = kx - \omega(k)t. \quad (2.23)$$

We wish to understand what are the values of x and t for which the packet $\psi(x, t)$ takes large values. We use the *stationary phase principle*: since only for $k \sim k_0$ the integral over k has a chance to give a non-zero contribution, the phase factor must be *stationary* at $k = k_0$. The idea is simple: if a function is multiplied by a rapidly varying phase, the integral washes out. Thus the phase must have zero derivative at k_0 . Applying this idea to our phase we find the derivative and set it equal to zero at k_0 :

$$\frac{d\varphi}{dk} \Big|_{k_0} = x - \frac{d\omega}{dk} \Big|_{k_0} t = 0. \quad (2.24)$$

This means that $\psi(x, t)$ is appreciable when x and t are related by

$$x = \frac{d\omega}{dk} \Big|_{k_0} t, \quad (2.25)$$

showing that the packet moves with *group velocity*

$$v_g = \frac{d\omega}{dk} \Big|_{k_0}. \quad (2.26)$$

Exercise. If $\Phi(k_0)$ is not real write $\Phi(k) = |\Phi(k)|e^{i\phi(k)}$. Find the new version of (2.25) and show that the velocity of the wave is not changed.

Let us now do a more detailed calculation that confirms the above analysis and gives some extra insight. Notice first that

$$\psi(x, 0) = \int dk \Phi(k) e^{ikx}. \quad (2.27)$$

We expand $\omega(k)$ in a Taylor expansion around $k = k_0$

$$\omega(k) = \omega(k_0) + (k - k_0) \frac{d\omega}{dk} \Big|_{k_0} + \mathcal{O}((k - k_0)^2). \quad (2.28)$$

Then we find, neglecting the $\mathcal{O}((k - k_0)^2)$ terms

$$\psi(x, t) = \int dk \Phi(k) e^{ikx} e^{-i\omega(k_0)t} e^{-i(k-k_0)\frac{d\omega}{dk}|_{k_0} t}. \quad (2.29)$$

It is convenient to take out of the integral all the factors that do not depend on k :

$$\begin{aligned} \psi(x, t) &= e^{-i\omega(k_0)t + ik_0 \frac{d\omega}{dk}|_{k_0} t} \int dk \Phi(k) e^{ikx} e^{-ik \frac{d\omega}{dk}|_{k_0} t} \\ &= e^{-i\omega(k_0)t + ik_0 \frac{d\omega}{dk}|_{k_0} t} \int dk \Phi(k) e^{ik(x - \frac{d\omega}{dk}|_{k_0} t)}. \end{aligned} \quad (2.30)$$

Comparing with (2.27) we realize that the integral in the above expression can be written in terms of the wavefunction at zero time:

$$\psi(x, t) = e^{-i\omega(k_0)t + ik_0 \frac{d\omega}{dk}|_{k_0} t} \psi\left(x - \frac{d\omega}{dk} \Big|_{k_0} t\right). \quad (2.31)$$

The phase factors in front of the expression are not important in tracking where the wave packet is. In particular we can take the norm of both sides of the equation to find

$$|\psi(x, t)| = \left| \psi\left(x - \frac{d\omega}{dk} \Big|_{k_0} t, 0\right) \right|. \quad (2.32)$$

If $\psi(x, 0)$ peaks at some value x_0 it is clear from the above equation that $|\psi(x, t)|$ peaks for

$$x - \frac{d\omega}{dk} \Big|_{k_0} t = x_0 \rightarrow x = x_0 + \frac{d\omega}{dk} \Big|_{k_0} t, \quad (2.33)$$

showing that the peak of the packet moves with velocity $v_{\text{gr}} = \frac{d\omega}{dk}$, evaluated at k_0 .

3 Choosing the wavefunction for a free particle

What is the mathematical form of the wave associated with a particle a particle with energy E and momentum p ? We know that ω and k are determined from $E = \hbar\omega$ and $p = \hbar k$. Let's suppose that we want our wave to be propagating in the $+\hat{x}$ direction. All the following are examples of waves that could be candidates for the particle wavefunction.

1. $\sin(kx - \omega t)$
2. $\cos(kx - \omega t)$
3. $e^{i(kx - \omega t)} = e^{ikx} e^{-i\omega t}$ - time dependence $\propto e^{-i\omega t}$
4. $e^{-i(kx - \omega t)} = e^{-ikx} e^{i\omega t}$ - time dependence $\propto e^{+i\omega t}$

In the third and fourth options we have indicated that the time dependence could come with either sign. We will use superposition to decide which is the right one! We are looking for a wave-function which is non-zero for all values of x .

Let's take them one by one:

1. Starting from (1), we build a superposition in which the particle has equal probability to be found moving in the $+x$ and the $-x$ directions.

$$\Psi(x, t) = \sin(kx - \omega t) + \sin(kx + \omega t) \quad (3.1)$$

Expanding the trigonometric functions this can be simplified to

$$\Psi(x, t) = 2 \sin(kx) \cos(\omega t). \quad (3.2)$$

But this result is not sensible. The wave function vanishes identically for all x at some special times

$$\omega t = \left(\frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \dots \right) \quad (3.3)$$

A wavefunction that is zero cannot represent a particle.

2. Constructing a wave function from (2) with a superposition of left and right going cos waves,

$$\Psi(x, t) = \cos(kx - \omega t) + \cos(kx + \omega t) = 2 \cos(kx) \cos(\omega t). \quad (3.4)$$

This choice is no good, it also vanishes identically when $\omega t = \left(\frac{\pi}{2}, \frac{3\pi}{2}, \dots \right)$

3. Let's try a similar superposition of exponentials from (3), with both having the same time dependence

$$\Psi(x, t) = e^{i(kx - \omega t)} + e^{i(-kx - \omega t)} \quad (3.5)$$

$$= (e^{ikx} + e^{-ikx}) e^{-i\omega t} \quad (3.6)$$

$$= 2 \cos(kx) e^{-i\omega t}. \quad (3.7)$$

This wavefunction meets our criteria! It is never zero for all values of x because $e^{-i\omega t}$ is never zero.

4. A superposition of exponentials from (4) also meets our criteria

$$\Psi(x, t) = e^{-i(kx - \omega t)} + e^{-i(-kx - \omega t)} \quad (3.8)$$

$$= (e^{ikx} + e^{-ikx}) e^{i\omega t} \quad (3.9)$$

$$= 2 \cos kx e^{i\omega t}. \quad (3.10)$$

This is never zero for all values of x

Since both options (3) and (4) seem to work we ask: Can we use *both* (3) and (4) to represent a particle moving to the right (in the $+\hat{x}$ direction)? Let's assume that we can. Then, since adding a state to itself should not change the state, we could represent the right moving particle by using the sum of (3) and (4)

$$\Psi(x, t) = e^{i(kx - \omega t)} + e^{-i(kx - \omega t)} = 2 \cos(kx - \omega t). \quad (3.11)$$

This, however, is the same as (2), which we already showed leads to difficulties. Therefore we must choose between (3) and (4).

The choice is a matter of convention, and all physicists use the same convention. We take the free particle wavefunction to be

$\text{Free particle wavefunction : } \Psi(x, t) = e^{i(kx - \omega t)},$

(3.12)

representing a particle with

$$p = \hbar k, \quad \text{and} \quad E = \hbar \omega. \quad (3.13)$$

In three dimensions the corresponding wavefunction would be

$\text{Free particle wavefunction : } \Psi(\mathbf{x}, t) = e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)},$

(3.14)

representing a particle with

$$\mathbf{p} = \hbar \mathbf{k}, \quad \text{and} \quad E = \hbar \omega. \quad (3.15)$$

Andrew Turner and Sarah Geller transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare
<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

Lecture 14 and 15: Algebraic approach to the SHO

B. Zwiebach
April 5, 2016

Contents

1 Algebraic Solution of the Oscillator	1
2 Operator manipulation and the spectrum	4

1 Algebraic Solution of the Oscillator

We have already seen how to calculate energy eigenstates for the simple harmonic oscillator by solving a second-order differential equation, the time-independent Schrödinger equation.

Let us now try to factorize the harmonic oscillator Hamiltonian. By this we mean, roughly, writing the Hamiltonian as the product of an operator times its *Hermitian conjugate*. As a first step we rewrite the Hamiltonian as

$$\hat{H} = \frac{1}{2}m\omega^2 \left(\hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2} \right). \quad (1.1)$$

Motivated by the identity $a^2 + b^2 = (a - ib)(a + ib)$, holding for numbers a and b , we examine if the expression in parenthesis can be written as a product

$$\begin{aligned} \left(\hat{x} - \frac{i\hat{p}}{m\omega} \right) \left(\hat{x} + \frac{i\hat{p}}{m\omega} \right) &= \hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2} + \frac{i}{m\omega} (\hat{x}\hat{p} - \hat{p}\hat{x}), \\ &= \hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2} - \frac{\hbar}{m\omega} \mathbf{1}, \end{aligned} \quad (1.2)$$

where the extra terms arise because \hat{x} and \hat{p} , as opposed to numbers, do not commute. We now define the right-most factor in the above product to be V :

$$V \equiv \hat{x} + \frac{i\hat{p}}{m\omega}, \quad (1.3)$$

Since \hat{x} and \hat{p} are Hermitian operators, we then have

$$V^\dagger = \hat{x} - \frac{i\hat{p}}{m\omega}, \quad (1.4)$$

and this is the left-most factor in the product! We can therefore rewrite (1.2) as

$$\hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2} = V^\dagger V + \frac{\hbar}{m\omega} \mathbf{1}, \quad (1.5)$$

and therefore back in the Hamiltonian (1.1) we find,

$$\hat{H} = \frac{1}{2}m\omega^2 V^\dagger V + \frac{1}{2}\hbar\omega \mathbf{1}. \quad (1.6)$$

This is a factorized form of the Hamiltonian: up to an additive constant E_0 , \hat{H} is the product of a positive constant times the operator product $V^\dagger V$. We note that the commutator of V and V^\dagger is simple

$$[V, V^\dagger] = \left[\hat{x} + \frac{i\hat{p}}{m\omega}, \hat{x} - \frac{i\hat{p}}{m\omega} \right] = -\frac{i}{m\omega}[\hat{x}, \hat{p}] + \frac{i}{m\omega}[\hat{p}, \hat{x}] = \frac{2\hbar}{m\omega}\mathbf{1}. \quad (1.7)$$

This implies that

$$\left[\sqrt{\frac{m\omega}{2\hbar}}V, \sqrt{\frac{m\omega}{2\hbar}}V^\dagger \right] = \mathbf{1}. \quad (1.8)$$

This suggests the definition of unit-free operator operators \hat{a} and \hat{a}^\dagger :

$$\begin{aligned} \hat{a} &\equiv \sqrt{\frac{m\omega}{2\hbar}}V, \\ \hat{a}^\dagger &\equiv \sqrt{\frac{m\omega}{2\hbar}}V^\dagger. \end{aligned} \quad (1.9)$$

Due to the scaling we have

$$[\hat{a}, \hat{a}^\dagger] = \mathbf{1}. \quad (1.10)$$

The operator \hat{a} is called *annihilation* operator and \hat{a}^\dagger is called a *creation* operator. The justification for these names will be seen below. From the above definitions we read the relations between $(\hat{a}, \hat{a}^\dagger)$ and (\hat{x}, \hat{p}) :

$$\boxed{\begin{aligned} \hat{a} &= \sqrt{\frac{m\omega}{2\hbar}} \left(\hat{x} + \frac{i\hat{p}}{m\omega} \right), \\ \hat{a}^\dagger &= \sqrt{\frac{m\omega}{2\hbar}} \left(\hat{x} - \frac{i\hat{p}}{m\omega} \right). \end{aligned}} \quad (1.11)$$

The inverse relations are many times useful as well,

$$\boxed{\begin{aligned} \hat{x} &= \sqrt{\frac{\hbar}{2m\omega}}(\hat{a} + \hat{a}^\dagger), \\ \hat{p} &= i\sqrt{\frac{m\omega\hbar}{2}}(\hat{a}^\dagger - \hat{a}). \end{aligned}} \quad (1.12)$$

While neither \hat{a} nor \hat{a}^\dagger is hermitian (they are hermitian conjugates of each other), the above equations are consistent with the hermiticity of \hat{x} and \hat{p} . We can now write the Hamiltonian in terms of the \hat{a} and \hat{a}^\dagger operators. Using (1.9) we have

$$V^\dagger V = \frac{2\hbar}{m\omega} \hat{a}^\dagger \hat{a}, \quad (1.13)$$

and therefore back in (1.6) we get

$$\boxed{\hat{H} = \hbar\omega(\hat{a}^\dagger \hat{a} + \frac{1}{2}) = \hbar\omega(\hat{N} + \frac{1}{2}), \quad \hat{N} \equiv \hat{a}^\dagger \hat{a}.} \quad (1.14)$$

The above form of the Hamiltonian is factorized: up to an additive constant \hat{H} is the product of a positive constant times the operator product $\hat{a}^\dagger \hat{a}$. In here we have dropped the identity operator, which

is usually understood. We have also introduced the *number* operator \hat{N} . This is, by construction, a hermitian operator and it is, up to a scale and an additive constant, equal to the Hamiltonian. An eigenstate of \hat{H} is also an eigenstate of \hat{N} and it follows from the above relation that the respective eigenvalues E and N are related by

$$E = \hbar\omega \left(N + \frac{1}{2} \right). \quad (1.15)$$

Let us now show the powerful conclusions that arise from the factorized Hamiltonian. On any state ψ that is normalized we have

$$\langle \hat{H} \rangle_\psi = (\psi, \hat{H}\psi) = \hbar\omega(\psi, \hat{a}^\dagger \hat{a}\psi) + \frac{1}{2}\hbar\omega(\psi, \psi), \quad (1.16)$$

and moving the \hat{a}^\dagger to the first input, we get

$$\langle \hat{H} \rangle_\psi = \hbar\omega(\hat{a}\psi, \hat{a}\psi) + \frac{1}{2}\hbar\omega \geq \frac{1}{2}\hbar\omega. \quad (1.17)$$

The inequality follows because any expression of the form (φ, φ) is greater than or equal to zero. This shows that for any energy eigenstate with energy E : $\hat{H}\psi = E\psi$ we have

$$\text{Energy eigenstates: } E \geq \frac{1}{2}\hbar\omega. \quad (1.18)$$

This important result about the spectrum followed directly from the factorization of the Hamiltonian. But we also get the information required to find the ground state wavefunction. The minimum energy $\frac{1}{2}\hbar\omega$ will be realized for a state ψ if the term $(\hat{a}\psi, \hat{a}\psi)$ in (1.17) vanishes. For this to vanish $\hat{a}\psi$ must vanish. Therefore, the ground state wavefunction φ_0 must satisfy

$$\hat{a}\varphi_0 = 0. \quad (1.19)$$

The operator \hat{a} annihilates the ground state and this why \hat{a} is called the annihilation operator. Using the definition of \hat{a} in (1.11) and the position space representation of \hat{p} , this becomes

$$\left(x + \frac{i}{m\omega} \frac{\hbar}{i} \frac{d}{dx} \right) \varphi_0(x) = 0 \rightarrow \left(x + \frac{\hbar}{m\omega} \frac{d}{dx} \right) \varphi_0(x) = 0. \quad (1.20)$$

Remarkably, this is a **first order** differential equation for the ground state. Not a second order equation, like the Schrödinger equation that determines the general energy eigenstates. This is a dramatic simplification afforded by the factorization of the Hamiltonian into a product of first-order differential operators. The above equation is rearranged as

$$\frac{d\varphi_0}{dx} = -\frac{m\omega}{\hbar} x \varphi_0. \quad (1.21)$$

Solving this differential equation yields

$$\varphi_0(x) = \left(\frac{m\omega}{\pi\hbar} \right)^{\frac{1}{4}} e^{-\frac{m\omega}{2\hbar}x^2},$$

(1.22)

where we included a normalization constant to guarantee that $(\varphi_0, \varphi_0) = 1$. Note that φ_0 is indeed an energy eigenstate with energy E_0 :

$$\hat{H}\varphi_0 = \hbar\omega(\hat{a}^\dagger \hat{a} + \frac{1}{2})\varphi_0 = \frac{1}{2}\hbar\omega\varphi_0 \rightarrow E_0 = \frac{1}{2}\hbar\omega. \quad (1.23)$$

Before proceeding with the analysis of excited states, let us view the properties of factorization more generally. Factorizing a Hamiltonian means finding an operator \hat{A} such that we can rewrite the Hamiltonian as $\hat{A}^\dagger \hat{A}$ up to an additive constant. Here \hat{A}^\dagger is the Hermitian conjugate of \hat{A} , an operator that is defined by

$$(\psi, \hat{A}^\dagger \varphi) = (\hat{A}\psi, \varphi). \quad (1.24)$$

We say that we have factorized a Hamiltonian \hat{H} if we can find a \hat{A} for which

$$\boxed{\hat{H} = \hat{A}^\dagger \hat{A} + E_0 \mathbf{1}}, \quad (1.25)$$

where E_0 is a constant with units of energy that multiplies the identity operator. This constant does not complicate our task of finding the eigenstates of the Hamiltonian, nor their energies: any eigenfunction of $\hat{A}^\dagger \hat{A}$ is an eigenfunction of \hat{H} . Two key properties follow from the factorization (1.25).

1. Any energy eigenstate must have energy greater than or equal to E_0 . First note that for an *arbitrary* normalized $\psi(x)$ we have

$$(\psi, \hat{H}\psi) = (\psi, \hat{A}^\dagger \hat{A}\psi) + E_0(\psi, \psi) = (\hat{A}\psi, \hat{A}\psi) + E_0, \quad (1.26)$$

Since the overlap $(\hat{A}\psi, \hat{A}\psi)$ is greater than or equal to zero, we have shown that

$$\boxed{(\psi, \hat{H}\psi) \geq E_0}. \quad (1.27)$$

If we take ψ to be an energy eigenstate of energy E : $\hat{H}\psi = E\psi$, the above relation gives

$$E \geq E_0. \quad (1.28)$$

This shows, as claimed, that all possible energies are greater than or equal to E_0 .

2. A wavefunction ψ_0 that satisfies

$$\hat{A}\psi_0 = 0, \quad (1.29)$$

is an energy eigenstate that *saturates* the inequality (1.28). Indeed,

$$\hat{H}\psi_0 = \hat{A}^\dagger \hat{A}\psi_0 + E_0\psi_0 = \hat{A}^\dagger(\hat{A}\psi_0) + E_0\psi_0 = E_0\psi_0. \quad (1.30)$$

The state ψ_0 satisfying $\hat{A}\psi_0 = 0$ is the ground state. For conventional Hamiltonians this is a first order differential equation for ψ_0 and much easier to solve than the Schrödinger equation.

2 Operator manipulation and the spectrum

We have seen that all energy eigenstates are eigenstates of the Hermitian number operator $\hat{N} = \hat{a}^\dagger \hat{a}$. This is because $\hat{H} = \hbar\omega(\hat{N} + \frac{1}{2})$. Note that since $\hat{a}\varphi_0 = 0$ we also have

$$\hat{N}\varphi_0 = 0. \quad (2.1)$$

We can quickly check that

$$\begin{aligned} [\hat{N}, \hat{a}] &= [\hat{a}^\dagger \hat{a}, \hat{a}] = [\hat{a}^\dagger, \hat{a}] \hat{a} = -\hat{a}, \\ [\hat{N}, \hat{a}^\dagger] &= [\hat{a}^\dagger \hat{a}, \hat{a}^\dagger] = \hat{a}^\dagger [\hat{a}, \hat{a}^\dagger] = \hat{a}^\dagger, \end{aligned} \quad (2.2)$$

which we summarize as

$$\begin{aligned} [\hat{N}, \hat{a}] &= -\hat{a}, \\ [\hat{N}, \hat{a}^\dagger] &= \hat{a}^\dagger. \end{aligned} \quad (2.3)$$

Using these identities and induction you should be able to show that:

$$\begin{aligned} [\hat{N}, (\hat{a})^k] &= -k(\hat{a})^k, \\ [\hat{N}, (\hat{a}^\dagger)^k] &= k(\hat{a}^\dagger)^k. \end{aligned} \quad (2.4)$$

These relations suggest why \hat{N} is called the number operator. Acting on powers of creation or annihilation operators by commutation it gives the same object multiplied by (plus or minus) the number of creation or annihilation operators, k in the above. Closely related commutators are also useful:

$$\begin{aligned} [\hat{a}^\dagger, (\hat{a})^k] &= -k(\hat{a})^{k-1} \\ [\hat{a}, (\hat{a}^\dagger)^k] &= k(\hat{a}^\dagger)^{k-1}. \end{aligned} \quad (2.5)$$

These commutators are analogous to $[\hat{p}, (\hat{x})^k]$ and $[\hat{x}, (\hat{p})^k]$. We will also make use of the following Lemma which helps in evaluations where we have an operator \hat{A} that kills a state ψ and we aim to simplify the action of $\hat{A}\hat{B}$, where \hat{B} is another operator, acting on ψ . Here is the result

If $\hat{A}\psi = 0$, then $\hat{A}\hat{B}\psi = [\hat{A}, \hat{B}]\psi$.

(2.6)

This is easily proved. First note that

$$\hat{A}\hat{B} = [\hat{A}, \hat{B}] + \hat{B}\hat{A}, \quad (2.7)$$

as can be quickly checked expanding the right-hand side. It then follows that

$$\hat{A}\hat{B}\psi = ([\hat{A}, \hat{B}] + \hat{B}\hat{A})\psi = [\hat{A}, \hat{B}]\psi, \quad (2.8)$$

because $\hat{B}\hat{A}\psi = \hat{B}(\hat{A}\psi) = 0$. This is what we wanted to show. This is all we need to know about commutators and we can now proceed to construct the states of the harmonic oscillator.

Since \hat{a} annihilates φ_0 consider acting on the ground state with \hat{a}^\dagger . It is clear that \hat{a}^\dagger cannot also annihilate φ_0 . If that would happen acting with both sides of the commutator identity $[\hat{a}, \hat{a}^\dagger] = 1$ on φ_0 would lead to a contradiction: the left-hand side would vanish but the right-hand side would not. Thus consider the wavefunction

$$\varphi_1 \equiv \hat{a}^\dagger \varphi_0. \quad (2.9)$$

We are going to show that this is an energy eigenstate. For this purpose we act on it with the number operator:

$$\hat{N}\varphi_1 = \hat{N}\hat{a}^\dagger \varphi_0 = [\hat{N}, \hat{a}^\dagger]\varphi_0, \quad (2.10)$$

where we noted that $\hat{N}\varphi_0 = 0$ and used Lemma (2.6). Given that $[\hat{N}, \hat{a}^\dagger] = \hat{a}^\dagger$, we get

$$\hat{N}\varphi_1 = \hat{a}^\dagger\varphi_0 = \varphi_1. \quad (2.11)$$

Thus φ_1 is an eigenstate of the operator \hat{N} with eigenvalue $N = 1$. Since φ_0 has \hat{N} eigenvalue zero, the effect of acting on φ_0 with \hat{a}^\dagger was to increase the eigenvalue of the number operator by one unit. The operator \hat{a}^\dagger is called the *creation* operator because it creates a state out of the ground state. Alternatively, it is called the *raising* operator, because it raises (by one unit) the eigenvalue of \hat{N} . Since $N = 1$ for φ_1 it follows that φ_1 is an energy eigenstate with energy E_1 given by

$$E_1 = \hbar\omega(1 + \frac{1}{2}) = \frac{3}{2}\hbar\omega. \quad (2.12)$$

It also turns out that φ_1 is properly normalized:

$$(\varphi_1, \varphi_1) = (\hat{a}^\dagger\varphi_0, \hat{a}^\dagger\varphi_0) = (\varphi_0, \hat{a}\hat{a}^\dagger\varphi_0), \quad (2.13)$$

where we used the Hermitian conjugation property to move the \hat{a}^\dagger acting on the left input into the right input, where it goes as $(\hat{a}^\dagger)^\dagger = \hat{a}$. We then have

$$(\varphi_1, \varphi_1) = (\varphi_0, \hat{a}\hat{a}^\dagger\varphi_0) = (\varphi_0, [\hat{a}, \hat{a}^\dagger]\varphi_0) = (\varphi_0, \varphi_0) = 1, \quad (2.14)$$

where we used (2.6) in the evaluation of $\hat{a}\hat{a}^\dagger\psi_0$. Indeed the state φ_1 is correctly normalized.

Next consider the state

$$\varphi'_2 \equiv \hat{a}^\dagger\hat{a}^\dagger\varphi_0. \quad (2.15)$$

This has

$$\hat{N}\varphi'_2 = \hat{N}\hat{a}^\dagger\hat{a}^\dagger\varphi_0 = [\hat{N}, \hat{a}^\dagger\hat{a}^\dagger]\varphi_0 = 2\hat{a}^\dagger\hat{a}^\dagger\varphi_0 = 2\varphi'_2, \quad (2.16)$$

so φ_2 is a state with number $N = 2$ and energy $E_2 = \frac{5}{2}\hbar\omega$. Is it properly normalized? We find

$$\begin{aligned} (\varphi_2, \varphi_2) &= (\hat{a}^\dagger\hat{a}^\dagger\varphi_0, \hat{a}^\dagger\hat{a}^\dagger\varphi_0) = (\varphi_0, \hat{a}\hat{a}\hat{a}^\dagger\hat{a}^\dagger\varphi_0) = (\varphi_0, \hat{a}[\hat{a}, \hat{a}^\dagger]\hat{a}^\dagger\varphi_0) \\ &= (\varphi_0, 2\hat{a}\hat{a}^\dagger\varphi_0) = (\varphi_0, \varphi_0) = 2. \end{aligned} \quad (2.17)$$

The properly normalized wavefunction is therefore

$$\varphi_2 \equiv \frac{1}{\sqrt{2}}\hat{a}^\dagger\hat{a}^\dagger\varphi_0. \quad (2.18)$$

We now claim that the n -th excited state of the simple harmonic oscillator is

$$\varphi_n \equiv \frac{1}{\sqrt{n!}} \underbrace{\hat{a}^\dagger \cdots \hat{a}^\dagger}_n \varphi_0 = \frac{1}{\sqrt{n!}} (\hat{a}^\dagger)^n \varphi_0. \quad (2.19)$$

Exercise: Verify that this state has \hat{N} eigenvalue n .

Exercise: Verify that the state φ_n is properly normalized.

Since the \hat{N} eigenvalue of φ is n , its energy E_n is given by

$$E_n = \hbar\omega(n + \frac{1}{2}). \quad (2.20)$$

Since the various states φ_n are eigenstates of a Hermitian operator (the Hamiltonian \hat{H}) with different eigenvalues, they are orthonormal

$$(\varphi_n, \varphi_m) = \delta_{m,n}. \quad (2.21)$$

We now note that $\hat{a}\varphi_n$ is a state with $n - 1$ operators \hat{a}^\dagger acting on φ_0 because the \hat{a} eliminates one of the creation operators in φ_n . Thus we expect $\hat{a}\varphi_n \sim \varphi_{n-1}$. We can make this precise

$$\hat{a}\varphi_n = \hat{a}\frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n\varphi_0 = \frac{1}{\sqrt{n!}}[\hat{a}, (\hat{a}^\dagger)^n]\varphi_0 = \frac{n}{\sqrt{n!}}(\hat{a}^\dagger)^{n-1}\varphi_0. \quad (2.22)$$

At this point we use (2.19) with n set equal to $n - 1$ and thus get

$$\hat{a}\varphi_n = \frac{n}{\sqrt{n!}}\sqrt{(n-1)!}\varphi_{n-1} = \sqrt{n}\varphi_{n-1}. \quad (2.23)$$

By the action of \hat{a}^\dagger on φ_n we get

$$\hat{a}^\dagger\varphi_n = \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^{n+1}\varphi_0 = \frac{1}{\sqrt{n!}}\sqrt{(n+1)!}\varphi_{n+1} = \sqrt{n+1}\varphi_{n+1}. \quad (2.24)$$

Collecting the results, we have

$$\begin{aligned} \hat{a}\varphi_n &= \sqrt{n}\varphi_{n-1}, \\ \hat{a}^\dagger\varphi_n &= \sqrt{n+1}\varphi_{n+1}. \end{aligned}$$

(2.25)

These relations make it clear that \hat{a} lowers the number of any energy eigenstate by one unit, except for the vacuum φ_0 which it kills. The raising operator \hat{a}^\dagger increases the number of any eigenstate by one unit.

Exercise: Calculate the uncertainty Δx of position in the n -th energy eigenstate.

Solution: By definition,

$$(\Delta x)_n^2 = \langle \hat{x}^2 \rangle_{\varphi_n} - \langle \hat{x} \rangle_{\varphi_n}^2. \quad (2.26)$$

The expectation value $\langle \hat{x} \rangle$ vanishes for any energy eigenstate since we are integrating x , which is odd, against $|\varphi_n(x)|^2$, which is always even. Still, it is instructive to see how this happens explicitly:

$$\langle \hat{x} \rangle_{\varphi_n} = (\varphi_n, \hat{x}\varphi_n) = \sqrt{\frac{\hbar}{2m\omega}}(\varphi_n, (\hat{a} + \hat{a}^\dagger)\varphi_n), \quad (2.27)$$

using the formula for \hat{x} in terms of \hat{a} and \hat{a}^\dagger . The above overlap vanishes because $\hat{a}\varphi_n \sim \varphi_{n-1}$ and $\hat{a}^\dagger\varphi_n \sim \varphi_{n+1}$ and both φ_{n-1} and φ_{n+1} are orthogonal to φ_n . Now we compute the expectation value of \hat{x}^2

$$\begin{aligned} \langle \hat{x}^2 \rangle_{\varphi_n} &= (\varphi_n, \hat{x}^2\varphi_n) = \frac{\hbar}{2m\omega}(\varphi_n, (\hat{a} + \hat{a}^\dagger)(\hat{a} + \hat{a}^\dagger)\varphi_n) \\ &= \frac{\hbar}{2m\omega}(\varphi_n, (\hat{a}\hat{a} + \hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a} + \hat{a}^\dagger\hat{a}^\dagger)\varphi_n). \end{aligned} \quad (2.28)$$

Since $\hat{a}\hat{a}\varphi_n \sim \varphi_{n-2}$ and $\hat{a}^\dagger\hat{a}^\dagger\varphi_n \sim \varphi_{n+2}$ and both φ_{n-2} and φ_{n+2} are orthogonal to φ_n , the $\hat{a}\hat{a}$ and $\hat{a}^\dagger\hat{a}^\dagger$ terms do not contribute. We are left with

$$\langle \hat{x}^2 \rangle_{\varphi_n} = \frac{\hbar}{2m\omega}(\varphi_n, (\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a})\varphi_n). \quad (2.29)$$

At this point we recognize that $\hat{a}^\dagger \hat{a} = \hat{N}$ and that $\hat{a}\hat{a}^\dagger = [\hat{a}, \hat{a}^\dagger] + \hat{a}^\dagger \hat{a} = 1 + \hat{N}$. As a result

$$\langle \hat{x}^2 \rangle_{\varphi_n} = \frac{\hbar}{2m\omega} (\varphi_n, (1 + 2\hat{N})\varphi_n) = \frac{\hbar}{2m\omega} (1 + 2n). \quad (2.30)$$

We therefore have

$$(\Delta x)_n^2 = \frac{\hbar}{m\omega} \left(n + \frac{1}{2} \right). \quad (2.31)$$

The position uncertainty grows linearly with the number.

Sarah Geller and Andrew Turner transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare
<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

DIRAC's BRA AND KET NOTATION

B. Zwiebach

October 7, 2013

Contents

1 From inner products to bra-kets	1
2 Operators revisited	5
2.1 Projection Operators	6
2.2 Adjoint of a linear operator	8
2.3 Hermitian and Unitary Operators	9
3 Non-denumerable basis	11

1 From inner products to bra-kets

Dirac invented a useful alternative notation for inner products that leads to the concepts of bras and kets. The notation is sometimes more efficient than the conventional mathematical notation we have been using. It is also widely although not universally used. It all begins by writing the inner product differently. The rule is to turn inner products into bra-ket pairs as follows

$$\langle u, v \rangle \longrightarrow \langle u | v \rangle. \quad (1.1)$$

Instead of the inner product comma we simply put a vertical bar! We can translate our earlier discussion of inner products trivially. In order to make you familiar with the new look we do it.

We now write $\langle u | v \rangle = \langle v | u \rangle^*$, as well as $\langle v | v \rangle \geq 0$ for all v , while $\langle v | v \rangle = 0$ if and only if $v = 0$. We have linearity in the second argument

$$\langle u | c_1 v_1 + c_2 v_2 \rangle = c_1 \langle u | v_1 \rangle + c_2 \langle u | v_2 \rangle, \quad (1.2)$$

for complex constants c_1 and c_2 , but antilinearity in the first argument

$$\langle c_1 u_1 + c_2 u_2 | v \rangle = c_1^* \langle u_1 | v \rangle + c_2^* \langle u_2 | v \rangle. \quad (1.3)$$

Two vectors u and v for which $\langle u | v \rangle = 0$ are orthogonal. For the norm: $|v|^2 = \langle v | v \rangle$. The Schwarz inequality, for any pair u and v of vectors reads $|\langle u | v \rangle| \leq |u| |v|$.

For a given physical situation, the inner product must be defined and should satisfy the axioms. Let us consider two examples:

1. Let $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ and $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ be two vectors in a complex dimensional vector space of dimension two. We then define

$$\langle a|b\rangle \equiv a_1^*b_1 + a_2^*b_2. \quad (1.4)$$

You should confirm the axioms are satisfied.

2. Consider the complex vector space of complex function $f(x) \in \mathbb{C}$ with $x \in [0, L]$. Given two such functions $f(x), g(x)$ we define

$$\langle f|g\rangle \equiv \int_0^L f^*(x)g(x)dx. \quad (1.5)$$

The verification of the axioms is again quite straightforward.

A set of basis vectors $\{e_i\}$ labelled by the integers $i = 1, \dots, n$ satisfying

$$\langle e_i|e_j\rangle = \delta_{ij}, \quad (1.6)$$

is orthonormal. An arbitrary vector can be written as a linear superposition of basis states:

$$v = \sum_i \alpha_i e_i, \quad (1.7)$$

We then see that the coefficients are determined by the inner product

$$\langle e_k|v\rangle = \langle e_k|\sum_i \alpha_i e_i\rangle = \sum_i \alpha_i \langle e_k|e_i\rangle = \alpha_k. \quad (1.8)$$

We can therefore write

$$v = \sum_i e_i \langle e_i|v\rangle. \quad (1.9)$$

To obtain now bras and kets, we reinterpret the inner product. We want to “split” the inner product into two ingredients

$$\langle u|v\rangle \rightarrow \langle u| |v\rangle. \quad (1.10)$$

Here $|v\rangle$ is called a **ket** and $\langle u|$ is called a **bra**. We will view the ket $|v\rangle$ just as another way to represent the vector v . This is a small subtlety with the notation: we think of $v \in V$ as a vector and also $|v\rangle \in V$ as a vector. It is as if we added some decoration $| \rangle$ around the vector v to make it clear by inspection that it is a vector, perhaps like the usual top arrows that are added in some cases. The label in the ket is a vector and the ket itself is that vector!

Bras are somewhat different objects. We say that bras belong to the space V^* dual to V . Elements of V^* are linear maps from V to \mathbb{C} . In conventional mathematical notation one has a $v \in V$ and a linear function $\phi \in V^*$ such that $\phi(v)$, which denotes the action of the function of the vector v , is a number. In the bracket notation we have the replacements

$$\begin{aligned} v &\rightarrow |v\rangle, \\ \phi &\rightarrow \langle u|, \\ \phi_u(v) &\rightarrow \langle u|v\rangle, \end{aligned} \quad (1.11)$$

where we used the notation in (6.6). Our bras are labelled by vectors: the object inside the $\langle |$ is a vector. But bras are *not* vectors. If kets are viewed as column vectors, then bras are viewed as row vectors. In this way a bra to the left of a ket makes sense: matrix multiplication of a row vector times a column vector gives a number. Indeed, for vectors

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \quad (1.12)$$

we had

$$\langle a|b\rangle = a_1^*b_1 + a_2^*b_2 + \dots + a_n^*b_n \quad (1.13)$$

Now we think of this as

$$\langle a| = (a_1^*, a_2^*, \dots, a_n^*), \quad |b\rangle = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \quad (1.14)$$

and matrix multiplication gives us the desired answer

$$\langle a|b\rangle = (a_1^*, a_2^*, \dots, a_n^*) \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = a_1^*b_1 + a_2^*b_2 + \dots + a_n^*b_n. \quad (1.15)$$

Note that the bra labeled by the vector a is obtained by forming the row vector and complex conjugating the entries. More abstractly the bra $\langle u|$ labeled by the vector u is defined by its action on arbitrary vectors $|v\rangle$ as follows

$$\langle u| : |v\rangle \rightarrow \langle u|v\rangle. \quad (1.16)$$

As required by the definition, any linear map from V to \mathbb{C} defines a bra, and the corresponding underlying vector. For example let v be a generic vector:

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, \quad (1.17)$$

A linear map $f(v)$ that acting on a vector v gives a number is an expression of the form

$$f(v) = \alpha_1^*v_1 + \alpha_2^*v_2 + \dots + \alpha_n^*v_n. \quad (1.18)$$

It is a linear function of the components of the vector. The linear function is specified by the numbers α_i , and for convenience (and without loss of generality) we used their complex conjugates. Note that we need exactly n constants, so they can be used to assemble a row vector or a bra

$$\langle \alpha| = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*) \quad (1.19)$$

and the associated vector or ket

$$|\alpha\rangle = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \quad (1.20)$$

Note that, by construction

$$f(v) = \langle \alpha | v \rangle. \quad (1.21)$$

This illustrates the point that (i) bras represent dual objects that act on vectors and (ii) bras are labelled by vectors.

Bras can be added and can be multiplied by complex numbers and there is a zero bra defined to give zero acting on any vector, so V^* is also a complex vector space. As a bra, the linear superposition

$$\langle \omega | \equiv \alpha \langle a | + \beta \langle b | \in V^*, \quad \alpha, \beta \in \mathbb{C}, \quad (1.22)$$

is defined to act on a vector (ket) $|c\rangle$ to give the number

$$\alpha \langle a | c \rangle + \beta \langle b | c \rangle. \quad (1.23)$$

For any vector $|v\rangle \in V$ there is a *unique* bra $\langle v | \in V^*$. If there would be another bra $\langle v' |$ it would have to act on arbitrary vectors $|w\rangle$ just like $\langle v |$:

$$\langle v' | w \rangle = \langle v | w \rangle \rightarrow \langle w | v \rangle - \langle w | v' \rangle = 0 \rightarrow \langle w | v - v' \rangle = 0. \quad (1.24)$$

In the first step we used complex conjugation and in the second step linearity. Now the vector $v - v'$ must have zero inner product with *any* vector w , so $v - v' = 0$ and $v = v'$.

We can now reconsider equation (1.3) and write an extra right-hand side

$$\langle \alpha_1 a_1 + \alpha_2 a_2 | b \rangle = \alpha_1^* \langle a_1 | b \rangle + \alpha_2^* \langle a_2 | b \rangle = (\alpha_1^* \langle a_1 | + \alpha_2^* \langle a_2 |) |b\rangle \quad (1.25)$$

so that we conclude that the rules to pass from kets to bras include

$$|v\rangle = \alpha_1 |a_1\rangle + \alpha_2 |a_2\rangle \longleftrightarrow \langle v | = \alpha_1^* \langle a_1 | + \alpha_2^* \langle a_2 |. \quad (1.26)$$

For simplicity of notation we sometimes write kets with labels simpler than vectors. Let us reconsider the basis vectors $\{e_i\}$ discussed in (1.6). The ket $|e_i\rangle$ is simply called $|i\rangle$ and the orthonormal condition reads

$$\langle i | j \rangle = \delta_{ij}. \quad (1.27)$$

The expansion (1.7) of a vector now reads

$$|v\rangle = \sum_i |i\rangle \alpha_i, \quad (1.28)$$

As in (1.8) the expansion coefficients are $\alpha_k = \langle k | v \rangle$ so that

$$|v\rangle = \sum_i |i\rangle \langle i | v \rangle. \quad (1.29)$$

2 Operators revisited

Let T be an operator in a vector space V . This means that acting on vectors on V it gives vectors on V , something we write as

$$\Omega : V \rightarrow V. \quad (2.30)$$

We denote by $\Omega|a\rangle$ the vector obtained by acting with Ω on the vector $|a\rangle$:

$$|a\rangle \in V \rightarrow \Omega|a\rangle \in V. \quad (2.31)$$

The operator Ω is linear if additionally we have

$$\Omega(|a\rangle + |b\rangle) = \Omega|a\rangle + \Omega|b\rangle, \text{ and } \Omega(\alpha|a\rangle) = \alpha\Omega|a\rangle. \quad (2.32)$$

When kets are labeled by vectors we sometimes write

$$|\Omega a\rangle \equiv \Omega|a\rangle, \quad (2.33)$$

It is useful to note that a linear operator on V is also a linear operator on V^*

$$\Omega : V^* \rightarrow V^*, \quad (2.34)$$

We write this as

$$\langle a| \rightarrow \langle a|\Omega \in V^*. \quad (2.35)$$

The object $\langle a|\Omega$ is defined to be the bra that acting on the ket $|b\rangle$ gives the number $\langle a|\Omega|b\rangle$.

We can write operators in terms of bras and kets, written in a suitable order. As an example of an operator consider a bra $\langle a|$ and a ket $|b\rangle$. We claim that the object

$$\Omega = |a\rangle\langle b|, \quad (2.36)$$

is naturally viewed as a linear operator on V and on V^* . Indeed, acting on a vector we let it act as the bra-ket notation suggests:

$$\Omega|v\rangle \equiv |a\rangle\langle b|v\rangle \sim |a\rangle, \text{ since } \langle b|v\rangle \text{ is a number.} \quad (2.37)$$

Acting on a bra it gives a bra:

$$\langle w|\Omega \equiv \langle w|a\rangle\langle b| \sim \langle b|, \text{ since } \langle w|a\rangle \text{ is a number.} \quad (2.38)$$

Let us now review the description of operators as matrices. The choice of basis is ours to make. For simplicity, however, we will usually consider orthonormal bases.

Consider therefore, two vectors expanded in an orthonormal basis $\{|i\rangle\}$:

$$|a\rangle = \sum_n |n\rangle a_n, \quad |b\rangle = \sum_n |n\rangle b_n. \quad (2.39)$$

Assume $|b\rangle$ is obtained by the action of Ω on $|a\rangle$:

$$\Omega|a\rangle = |b\rangle \rightarrow \sum_n \Omega|n\rangle a_n = \sum_n |n\rangle b_n. \quad (2.40)$$

Acting on both sides of this vector equation with the bra $\langle m|$ we find

$$\sum_n \langle m|\Omega|n\rangle a_n = \sum_n \langle m|n\rangle b_n = b_m \quad (2.41)$$

We now define the ‘matrix elements’

$$\boxed{\Omega_{mn} \equiv \langle m|\Omega|n\rangle.} \quad (2.42)$$

so that the above equation reads

$$\sum_n \Omega_{mn} a_n = b_m, \quad (2.43)$$

which is the matrix version of the original relation $\Omega|a\rangle = |b\rangle$. The chosen basis has allowed us to view the linear operator Ω as a matrix, also denoted as Ω , with matrix components Ω_{mn} :

$$\Omega \longleftrightarrow \begin{pmatrix} \Omega_{11} & \Omega_{12} & \dots & \dots & \Omega_{1N} \\ \Omega_{21} & \Omega_{22} & \dots & \dots & \Omega_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Omega_{N1} & \Omega_{N2} & \dots & \dots & \Omega_{NN} \end{pmatrix}, \quad \text{with } \Omega_{ij} = \langle i|\Omega|j\rangle. \quad (2.44)$$

There is one additional claim. The operator itself can be written in terms of the matrix elements and basis bras and kets. We claim that

$$\boxed{\Omega = \sum_{m,n} |m\rangle \Omega_{mn} \langle n|.} \quad (2.45)$$

We can verify that this is correct by computing the matrix elements using it:

$$\langle m'|\Omega|n'\rangle = \sum_{m,n} \Omega_{mn} \langle m'|m\rangle \langle n|n'\rangle = \sum_{m,n} \Omega_{mn} \delta_{m'm} \delta_{nn'} = \Omega_{m'n'}, \quad (2.46)$$

as expected from the definition (2.42).

2.1 Projection Operators

Consider the familiar orthonormal basis $\{|i\rangle\}$ of V and choose one element $|m\rangle$ from the basis to form an operator P_m defined by

$$P_m \equiv |m\rangle \langle m|. \quad (2.47)$$

This operator maps any vector $|v\rangle \in V$ to a vector along $|, \rangle$. Indeed, acting on $|v\rangle$ it gives

$$P_m|v\rangle = |m\rangle \langle m|v\rangle \sim |m\rangle. \quad (2.48)$$

Comparing the above expression for P_m with (2.45) we see that in the chosen basis, P_n is represented by a matrix all of whose elements are zero, except for the (n, n) element $(P_n)_{nn}$ which is one:

$$P_n \longleftrightarrow \begin{pmatrix} 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \end{pmatrix}. \quad (2.49)$$

A hermitian operator P is said to be a *projection* operator if it satisfies the operator equation $PP = P$. This means that acting twice with a projection operator on a vector gives the same as acting once. The operator P_m is a projection operator since

$$P_m P_m = (|m\rangle\langle m|)(|m\rangle\langle m|) = |m\rangle\langle m|m\rangle\langle m| = |m\rangle\langle m|, \quad (2.50)$$

since $\langle m|m\rangle = 1$. The operator P_m is said to be a *rank one* projection operator since it projects to a one-dimensional subspace of V , the subspace generated by $|m\rangle$.

Using the basis vector $|m\rangle$ with $m \neq n$ we can define

$$P_{m,n} \equiv |m\rangle\langle m| + |n\rangle\langle n|. \quad (2.51)$$

Acting on any vector $|v\rangle \in V$, this operator gives us a vector in the subspace spanned by $|m\rangle$ and $|n\rangle$:

$$P_{m,n}|v\rangle = |m\rangle\langle m|v\rangle + |n\rangle\langle n|v\rangle. \quad (2.52)$$

Using the orthogonality of $|m\rangle$ and $|n\rangle$ we quickly find that $P_{m,n}P_{m,n} = P_{m,n}$ and therefore $P_{m,n}$ is a projector. It is a rank two projector, since it projects to a two-dimensional subspace of V , the subspace spanned by $|m\rangle$ and $|n\rangle$. Similarly, we can construct a rank three projector by adding an extra term $|k\rangle\langle k|$ with $k \neq m$ and $k \neq n$. If we include all basis vectors we would have the operator

$$P_{1,\dots,N} \equiv |1\rangle\langle 1| + |2\rangle\langle 2| + \dots + |N\rangle\langle N|. \quad (2.53)$$

As a matrix $P_{1,\dots,N}$ has a one on every element of the diagonal and a zero everywhere else. This is therefore the unit matrix, which represents the identity operator. Indeed we anticipated this in (1.29), and we thus write

$$\mathbf{1} = \sum_i |i\rangle\langle i|. \quad (2.54)$$

This is the completeness relation for the chosen orthonormal basis. This equation is sometimes called the ‘resolution’ of the identity.

Example. For the spin one-half system the unit operator can be written as a sum of two terms since the vector space is two dimensional. Using the orthonormal basis vectors $|+\rangle$ and $|-\rangle$ for spins along the positive and negative z directions, respectively, we have

$$\mathbf{1} = |+\rangle\langle +| + |-\rangle\langle -|. \quad (2.55)$$

Example. We can use the completeness relation to show that our formula (2.42) for matrix elements is consistent with matrix multiplication. Indeed for the product $\Omega_1\Omega_2$ of two operators we write

$$\begin{aligned} (\Omega_1\Omega_2)_{mn} &= \langle m|\Omega_1\Omega_2|n\rangle = \langle m|\Omega_1 \mathbf{1} \Omega_2|n\rangle \\ &= \langle m|\Omega_1 \left(\sum_{k=1}^N |k\rangle\langle k| \right) \Omega_2|n\rangle = \sum_{k=1}^N \langle m|\Omega_1|k\rangle\langle k|\Omega_2|n\rangle = \sum_{k=1}^N (\Omega_1)_{mk}(\Omega_2)_{kn}. \end{aligned} \quad (2.56)$$

This is the expected rule for the multiplication of the matrices corresponding to Ω_1 and Ω_2 .

2.2 Adjoint of a linear operator

A linear operator Ω on V is defined by its action on the vectors in V . We have noted that Ω can also be viewed as a linear operator on the dual space V^* . We defined the linear operator Ω^\dagger associated with Ω . In general Ω^\dagger by

$$\langle \Omega^\dagger u | v \rangle = \langle u | \Omega v \rangle \quad (2.57)$$

Flipping the order on the left-hand side we get

$$\langle v | \Omega^\dagger u \rangle^* = \langle u | \Omega v \rangle \quad (2.58)$$

Complex conjugating, and writing the operators more explicitly

$$\boxed{\langle v | \Omega^\dagger u \rangle = \langle u | \Omega | v \rangle^*, \quad \forall u, v.} \quad (2.59)$$

Flipping the two sides of (2.57) we also get

$$\langle v | \Omega^\dagger | u \rangle = \langle \Omega v | u \rangle \quad (2.60)$$

from which, taking the ket away, we learn that

$$\boxed{\langle v | \Omega^\dagger \equiv \langle \Omega v |.} \quad (2.61)$$

Another way to state the action of the operator Ω^\dagger is as follows. The linear operator Ω induces a map $|v\rangle \rightarrow |v'\rangle$ of vectors in V and, in fact, is defined by giving a complete list of these maps. The operator Ω^\dagger is defined as the one that induces the maps $\langle v | \rightarrow \langle v' |$ of the *corresponding* bras. Indeed,

$$\begin{aligned} |v'\rangle &= |\Omega v\rangle = \Omega |v\rangle, \\ \langle v'| &= \langle \Omega v | = \langle v | \Omega^\dagger \end{aligned} \quad (2.62)$$

The first line is just definitions. On the second line, the first equality is obtained by taking bras of the first equality on the first line. The second equality is just (2.61). We say it as

$$\boxed{\text{The bra associated with } \Omega |v\rangle \text{ is } \langle v | \Omega^\dagger.} \quad (2.63)$$

To see what hermiticity means at the level of matrix elements, we take u, v to be orthonormal basis vectors in (2.59)

$$\langle i|\Omega^\dagger|j\rangle = \langle j|\Omega|i\rangle^* \rightarrow (\Omega^\dagger)_{ij} = (\Omega_{ji})^*. \quad (2.64)$$

In matrix notation we have $\Omega^\dagger = (\Omega^t)^*$ where the superscript t denotes transposition.

Exercise. Show that $(\Omega_1\Omega_2)^\dagger = \Omega_2^\dagger\Omega_1^\dagger$ by taking matrix elements.

Exercise. Given an operator $\Omega = |a\rangle\langle b|$ for arbitrary vectors a, b , write a bra-ket expression for Ω^\dagger .

Solution: Acting with Ω on $|v\rangle$ and then taking the dual gives

$$\Omega|v\rangle = |a\rangle\langle b|v\rangle \rightarrow \langle v|\Omega^\dagger = \langle v|b\rangle\langle a|, \quad (2.65)$$

Since this equation is valid for any bra $\langle v|$ we read

$$\Omega^\dagger = |b\rangle\langle a|. \quad (2.66)$$

2.3 Hermitian and Unitary Operators

A linear operator Ω is said to be *hermitian* if it is equal to its adjoint:

$$\boxed{\text{Hermitian Operator: } \Omega^\dagger = \Omega.} \quad (2.67)$$

In quantum mechanics Hermitian operators are associated with observables. The eigenvalues of a Hermitian operator are the possible measured values of the observables. As we will show soon, the eigenvalues of a Hermitian operator are all real. An operator A is said to be *anti-hermitian* if $A^\dagger = -A$.

Exercise: Show that the commutator $[\Omega_1, \Omega_2]$ of two hermitian operators Ω_1 and Ω_2 is anti-hermitian.

There are a couple of equations that rewrite in useful ways the main property of Hermitian operators. Using $\Omega^\dagger = \Omega$ in (2.59) we find

$$\boxed{\text{If } \Omega \text{ is a Hermitian Operator: } \langle v|\Omega|u\rangle = \langle u|\Omega|v\rangle^*, \quad \forall u, v.} \quad (2.68)$$

It follows that the expectation value of a Hermitian operator in *any* state is real

$$\langle v|\Omega|v\rangle \text{ is real for any hermitian } \Omega. \quad (2.69)$$

Another neat form of the hermiticity condition is derived as follows:

$$\langle \Omega u|v\rangle = \langle u|\Omega^\dagger|v\rangle = \langle u|\Omega|v\rangle = \langle u|\Omega v\rangle, \quad (2.70)$$

so that all in all

$$\boxed{\text{Hermitian Operator: } \langle \Omega u|v\rangle = \langle u|\Omega v\rangle.} \quad (2.71)$$

In this expression we see that a hermitian operator moves freely from the bra to the ket (and viceversa).

Example: For wavefunction $f(x) \in \mathbb{C}$ we have written

$$\langle f|g \rangle = \int_{-\infty}^{\infty} (f(x))^* g(x) dx \quad (2.72)$$

For a Hermitian Ω we have $\langle \Omega f | g \rangle = \langle f | \Omega g \rangle$ or explicitly

$$\int_{-\infty}^{\infty} (\Omega f(x))^* g(x) dx = \int_{-\infty}^{\infty} (f(x))^* \Omega g(x) dx \quad (2.73)$$

Verify that the linear operator $\Omega = \frac{\hbar}{i} \frac{d}{dx}$ is hermitian when we restrict to functions that vanish at $\pm\infty$.

An operator U is said to be a unitary operator if U^\dagger is an inverse for U , that is, $U^\dagger U$ and UU^\dagger are both the identity operator:

$$U \text{ is a unitary operator: } U^\dagger U = UU^\dagger = \mathbf{1} \quad (2.74)$$

In finite dimensional vector spaces $U^\dagger U = \mathbf{1}$ implies $UU^\dagger = 1$, but this is not always the case for infinite dimensional vector spaces. A key property of unitary operators is that they preserve the norm of states. Indeed, assume that $|\psi'\rangle$ is obtained by the action of U on $|\psi\rangle$:

$$|\psi'\rangle = U|\psi\rangle \quad (2.75)$$

Taking the dual we have

$$\langle \psi' | = \langle \psi | U^\dagger, \quad (2.76)$$

and therefore

$$\langle \psi' | \psi' \rangle = \langle \psi | U^\dagger U | \psi \rangle = \langle \psi | \psi \rangle, \quad (2.77)$$

showing that $|\psi\rangle$ and $U|\psi\rangle$ are states with the same norm. More generally

$$\langle Ua | Ub \rangle = \langle a | U^\dagger U | b \rangle = \langle a | b \rangle. \quad (2.78)$$

Another important property of unitary operators is that acting on an orthonormal basis they give another orthonormal basis. To show this consider the orthonormal basis

$$|a_1\rangle, |a_2\rangle, \dots |a_N\rangle, \quad \langle a_i | a_j \rangle = \delta_{ij} \quad (2.79)$$

Acting with U we get

$$|Ua_1\rangle, |Ua_2\rangle, \dots |Ua_N\rangle, \quad (2.80)$$

To show that this is a basis we must prove that

$$\sum_i \beta_i |Ua_i\rangle = 0 \quad (2.81)$$

implies $\beta_i = 0$ for all i . Indeed, the above gives

$$\sum_i \beta_i |Ua_i\rangle = \sum_i \beta_i U|a_i\rangle = U \sum_i \beta_i |a_i\rangle = 0. \quad (2.82)$$

Acting with U^\dagger from the left we find that $\sum_i \beta_i |a_i\rangle = 0$ and, since the $|a_i\rangle$ form a basis, we get $\beta_i = 0$ for all i , as desired. The new basis is orthonormal because

$$\langle Ua_i|Ua_j\rangle = \langle a_i|U^\dagger U|a_j\rangle = \langle a_i|a_j\rangle = \delta_{ij}. \quad (2.83)$$

It follows from the above that the operator U can be written as

$$U = \sum_{i=1}^N |Ua_i\rangle\langle a_i|, \quad (2.84)$$

since

$$U|a_j\rangle = \sum_{i=1}^N |Ua_i\rangle\langle a_i|a_j\rangle = |Ua_j\rangle. \quad (2.85)$$

In fact for *any* unitary operator U in a vector space V there exist orthonormal bases $\{|a_i\rangle\}$ and $\{|b_i\rangle\}$ such that U can be written as

$$U = \sum_{i=1}^N |b_i\rangle\langle a_i|. \quad (2.86)$$

Indeed, this is just a rewriting of (2.84), with $|a_i\rangle$ any orthonormal basis and $|b_i\rangle = |Ua_i\rangle$.

Exercise: Verify that U in (2.86) satisfies $U^\dagger U = UU^\dagger = \mathbf{1}$.

Exercise: Prove that $\langle a_i|U|a_j\rangle = \langle b_i|U|b_j\rangle$.

3 Non-denumerable basis

In this section we describe the use of bras and kets for the position and momentum states of a particle moving on the real line $x \in \mathbb{R}$.

Let us begin with position. We will introduce position states $|x\rangle$ where the label x in the ket is the value of the position. Since x is a continuous variable and we position states $|x\rangle$ for all values of x to form a basis, we are dealing with an infinite basis that is not possible to label as $|1\rangle, |2\rangle, \dots$, it is a non-denumerable basis. So we have

$$\text{Basis states : } |x\rangle, \quad \forall x \in \mathbb{R}. \quad (3.87)$$

Basis states with different values of x are different vectors in the state space (a complex vector space, as always in quantum mechanics). Note here that the label on the ket is not a vector! So $|ax\rangle \neq a|x\rangle$, for any real $a \neq 1$. In particular $|-x\rangle \neq |x\rangle$ unless $x = 0$. For quantum mechanics in three dimensions, we have position states $|\vec{x}\rangle$. Here the label is a vector in a three-dimensional real vector space (our space!) while the ket is a vector in the infinite dimensional complex vector space of states of the theory.

Again something like $|\vec{x}_1 + \vec{x}_2\rangle$ has nothing to do with $|\vec{x}_1\rangle + |\vec{x}_2\rangle$. The $| \rangle$ enclosing the label of the position eigenstates plays a crucial role: it helps us see that object lives in an infinite dimensional complex vector space.

The inner product must be defined, so we will take

$$\langle x|y\rangle = \delta(x - y). \quad (3.88)$$

It follows that position states with different positions are orthogonal to each other. The norm of a position state is infinite: $\langle x|x\rangle = \delta(0) = \infty$, so these are not allowed states of particles. We visualize the state $|x\rangle$ as the state of a particle perfectly localized at x , but this is an idealization. We can easily construct normalizable states using superpositions of position states. We also have a completeness relation

$$\mathbf{1} = \int dx |x\rangle\langle x|. \quad (3.89)$$

This is consistent with our inner product above. Letting the above equation act on $|y\rangle$ we find an equality:

$$|y\rangle = \int dx |x\rangle\langle x|y\rangle = \int dx |x\rangle \delta(x - y) = |y\rangle. \quad (3.90)$$

The position operator \hat{x} is defined by its action on the position states. Not surprisingly we let

$$\hat{x}|x\rangle = x|x\rangle, \quad (3.91)$$

thus declaring that $|x\rangle$ are \hat{x} eigenstates with eigenvalue equal to the position x . We can also show that \hat{x} is a Hermitian operator by checking that \hat{x}^\dagger and \hat{x} have the same matrix elements:

$$\langle x_1|\hat{x}^\dagger|x_2\rangle = \langle x_2|\hat{x}|x_1\rangle^* = [x_1\delta(x_1 - x_2)]^* = x_2\delta(x_1 - x_2) = \langle x_1|\hat{x}|x_2\rangle. \quad (3.92)$$

We thus conclude that $\hat{x}^\dagger = \hat{x}$ and the bra associated with (3.91) is

$$\langle x|\hat{x} = x\langle x|. \quad (3.93)$$

Given the state $|\psi\rangle$ of a particle, we define the associated position-state wavefunction $\psi(x)$ by

$$\psi(x) \equiv \langle x|\psi\rangle \in \mathbb{C}. \quad (3.94)$$

This is sensible: $\langle x|\psi\rangle$ is a number that depends on the value of x , thus a function of x . We can now do a number of basic computations. First we write any state as a superposition of position eigenstates, by inserting $\mathbf{1}$ as in the completeness relation

$$|\psi\rangle = \mathbf{1}|\psi\rangle = \int dx |x\rangle\langle x|\psi\rangle = \int dx |x\rangle \psi(x). \quad (3.95)$$

As expected, $\psi(x)$ is the component of ψ along the state $|x\rangle$. Overlap of states can also be written in position space:

$$\langle\phi|\psi\rangle = \int dx \langle\phi|x\rangle\langle x|\psi\rangle = \int dx \phi^*(x)\psi(x). \quad (3.96)$$

Matrix elements involving \hat{x} are also easily evaluated

$$\langle \phi | \hat{x} | \psi \rangle = \langle \phi | \hat{x} \mathbf{1} | \psi \rangle = \int dx \langle \phi | \hat{x} | x \rangle \langle x | \psi \rangle = \int dx \langle \phi | x \rangle x \langle x | \psi \rangle = \int dx \phi^*(x) x \psi(x). \quad (3.97)$$

We now introduce momentum states $|p\rangle$ that are eigenstates of the momentum operator \hat{p} in complete analogy to the position states

$$\begin{aligned} \text{Basis states : } & |p\rangle, \quad \forall p \in \mathbb{R}. \\ \langle p' | p \rangle &= \delta(p - p'), \\ \mathbf{1} &= \int dp |p\rangle \langle p|, \\ \hat{p} |p\rangle &= p |p\rangle \end{aligned} \quad (3.98)$$

Just as for coordinate space we also have

$$\hat{p}^\dagger = \hat{p}, \quad \text{and} \quad \langle p | \hat{p} = p \langle p|. \quad (3.99)$$

In order to relate the two bases we need the value of the overlap $\langle x | p \rangle$. Since we interpret this as the wavefunction for a particle with momentum p we have from (6.39) of Chapter 1 that

$$\langle x | p \rangle = \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}. \quad (3.100)$$

The normalization was adjusted properly to be compatible with the completeness relations. Indeed, for example, consider the $\langle p' | p \rangle$ overlap and use the completeness in x to evaluate it

$$\langle p' | p \rangle = \int dx \langle p' | x \rangle \langle x | p \rangle = \frac{1}{2\pi\hbar} \int dx e^{i(p-p')x/\hbar} = \frac{1}{2\pi} \int du e^{i(p-p')u}, \quad (3.101)$$

where we let $u = x/\hbar$ in the last step. We claim that the last integral is precisely the integral representation of the delta function $\delta(p - p')$:

$$\frac{1}{2\pi} \int du e^{i(p-p')u} = \delta(p - p'). \quad (3.102)$$

This, then gives the correct value for the overlap $\langle p | p' \rangle$, as we claimed. The integral (3.102) can be justified using the fact that the functions

$$f_n(x) \equiv \frac{1}{\sqrt{L}} \exp\left(\frac{2\pi i n x}{L}\right), \quad (3.103)$$

form a complete orthonormal set of functions over the interval $x \in [-L/2, L/2]$. Completeness then means that

$$\sum_{n \in \mathbb{Z}} f_n^*(x) f_n(x') = \delta(x - x'). \quad (3.104)$$

We thus have

$$\sum_{n \in \mathbb{Z}} \frac{1}{L} \exp\left(2\pi i \frac{n}{L}(x - x')\right) = \delta(x - x'). \quad (3.105)$$

In the limit as L goes to infinity the above sum can be written as an integral since the exponential is a very slowly varying function of $n \in \mathbb{Z}$. Since $\Delta n = 1$ with $u = 2\pi n/L$ we have $\Delta u = 2\pi/L \ll 1$ and then

$$\sum_{n \in \mathbb{Z}} \frac{1}{L} \exp\left(2\pi i \frac{n}{L}(x - x')\right) = \sum_u \frac{\Delta u}{2\pi} \exp\left(i u(x - x')\right) \rightarrow \frac{1}{2\pi} \int du e^{iu(x-x')} , \quad (3.106)$$

and back in (3.105) we have justified (3.102).

We can now ask: What is $\langle p|\psi\rangle$? We compute

$$\langle p|\psi\rangle = \int dx \langle p|x\rangle \langle x|\psi\rangle = \frac{1}{\sqrt{2\pi\hbar}} \int dx e^{-ipx/\hbar} \psi(x) = \tilde{\psi}(p) , \quad (3.107)$$

which is the Fourier transform of $\psi(x)$, as defined in (6.41) of Chapter 1. Thus the Fourier transform of $\psi(x)$ is the wavefunction in the momentum representation.

It is useful to know how to evaluate $\langle x|\hat{p}|\psi\rangle$. We do it by inserting a complete set of momentum states:

$$\langle x|\hat{p}|\psi\rangle = \int dp \langle x|p\rangle \langle p|\hat{p}|\psi\rangle = \int dp (p\langle x|p\rangle) \langle p|\psi\rangle \quad (3.108)$$

Now we notice that

$$p\langle x|p\rangle = \frac{\hbar}{i} \frac{d}{dx} \langle x|p\rangle \quad (3.109)$$

and thus

$$\langle x|\hat{p}|\psi\rangle = \int dp \left(\frac{\hbar}{i} \frac{d}{dx} \langle x|p\rangle \right) \langle p|\psi\rangle . \quad (3.110)$$

The derivative can be moved out of the integral, since no other part of the integrand depends on x :

$$\langle x|\hat{p}|\psi\rangle = \frac{\hbar}{i} \frac{d}{dx} \int dp \langle x|p\rangle \langle p|\psi\rangle \quad (3.111)$$

The completeness sum is now trivial and can be discarded to obtain

$$\langle x|\hat{p}|\psi\rangle = \frac{\hbar}{i} \frac{d}{dx} \langle x|\psi\rangle = \frac{\hbar}{i} \frac{d}{dx} \psi(x) . \quad (3.112)$$

Exercise. Show that

$$\langle p|\hat{x}|\psi\rangle = i\hbar \frac{d}{dp} \tilde{\psi}(p) . \quad (3.113)$$

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

WAVE MECHANICS

B. Zwiebach
September 13, 2013

Contents

1 The Schrödinger equation	1
2 Stationary Solutions	4
3 Properties of energy eigenstates in one dimension	10
4 The nature of the spectrum	12
5 Variational Principle	18
6 Position and momentum	22

1 The Schrödinger equation

In classical mechanics the motion of a particle is usually described using the time-dependent position $\vec{x}(t)$ as the dynamical variable. In wave mechanics the dynamical variable is a wavefunction. This wavefunction depends on position and on time and it is a complex number – it belongs to the complex numbers \mathbb{C} (we denote the real numbers by \mathbb{R}). When all three dimensions of space are relevant we write the wavefunction as

$$\Psi(\vec{x}, t) \in \mathbb{C}. \quad (1.1)$$

When only one spatial dimension is relevant we write it as $\Psi(x, t) \in \mathbb{C}$. The wavefunction satisfies the Schrödinger equation. For one-dimensional space we write

$$i\hbar \frac{\partial \Psi}{\partial t}(x, t) = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x, t) \right) \Psi(x, t). \quad (1.2)$$

This is the equation for a (non-relativistic) particle of mass m moving along the x axis while acted by the potential $V(x, t) \in \mathbb{R}$. It is clear from this equation that the wavefunction must be complex: if it were real, the right-hand side of (1.2) would be real while the left-hand side would be imaginary, due to the explicit factor of i .

Let us make two important remarks:

1. The Schrödinger equation is a *first order* differential equation in time. This means that if we prescribe the wavefunction $\Psi(x, t_0)$ for all of space at an arbitrary initial time t_0 , the wavefunction is determined for all times.
2. The Schrödinger equation is a *linear* equation for Ψ : if Ψ_1 and Ψ_2 are solutions so is $a_1\Psi_1 + a_2\Psi_2$ with a_1 and a_2 arbitrary *complex numbers*.

Given a complex number $z = a + ib$, $a, b \in \mathbb{R}$, its complex conjugate is $z^* = a - ib$. Let $|z|$ denote the norm or length of the complex number z . The norm is a positive number (thus real!) and it is given by $|z| = \sqrt{a^2 + b^2}$. If the norm of a complex number is zero, the complex number is zero. You can quickly verify that

$$|z|^2 = zz^*. \quad (1.3)$$

For a wavefunction $\Psi(x, t)$ its complex conjugate $(\Psi(x, t))^*$ will be usually written as $\Psi^*(x, t)$.

We define the probability density $P(x, t)$, also denoted as $\rho(x, t)$, as the norm-squared of the wavefunction:

$$P(x, t) = \rho(x, t) \equiv \Psi^*(x, t)\Psi(x, t) = |\Psi(x, t)|^2. \quad (1.4)$$

This probability density so defined is positive. The physical interpretation of the wavefunction arises because we declare that

$P(x, t) dx$ is the probability to find the particle in the interval $[x, x + dx]$ at time t .

(1.5)

This interpretation requires a *normalized* wavefunction, namely, the wavefunction used above must satisfy, for all times,

$$\int_{-\infty}^{\infty} dx |\Psi(x, t)|^2 = 1, \quad \forall t. \quad (1.6)$$

By integrating over space, the left-hand adds up the probabilities that the particle be found in all of the tiny intervals dx that comprise the real line. Since the particle must be found somewhere this sum must be equal to one.

Suppose you are handed a wavefunction that is normalized at time t_0 :

$$\int_{-\infty}^{\infty} dx |\Psi(x, t_0)|^2 = 1, \quad \forall t. \quad (1.7)$$

As mentioned above, knowledge of the wavefunction at one time implies, via the Schrödinger equation, knowledge for all times. The Schrödinger equation must guarantee that the wavefunction remains normalized for all times. Proving this is a good exercise:

Exercise 1. Show that the Schrödinger equation implies that the norm of the wavefunction does not change in time:

$$\frac{d}{dt} \int_{-\infty}^{\infty} dx |\Psi(x, t)|^2 = 0. \quad (1.8)$$

You will have to use both the Schrödinger equation and its complex-conjugate version. Moreover you will have to use $\Psi(x, t) \rightarrow 0$ as $|x| \rightarrow \infty$, which is true, as no normalizable wavefunction can take a non-zero value as $|x| \rightarrow \infty$. While generally the derivative $\frac{\partial}{\partial x} \Psi$ also goes to zero as $|x| \rightarrow \infty$ you only need to assume that it remains bounded.

- - - - -

Associated to the probability density $\rho(x, t) = \Psi^* \Psi$ there is a **probability current** $J(x, t)$ that characterizes the flow of probability and is given by

$$J(x, t) = \frac{\hbar}{m} \text{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right). \quad (1.9)$$

The analogy in electromagnetism is useful. There we have the current density vector \vec{J} and the charge density ρ . The statement of charge conservation is the differential relation

$$\nabla \cdot \vec{J} + \frac{\partial \rho}{\partial t} = 0. \quad (1.10)$$

This equation applied to a fixed volume V implies that the rate of change of the enclosed charge $Q_V(t)$ is only due to the flux of \vec{J} across the surface S that bounds the volume:

$$\frac{dQ_V}{dt}(t) = - \oint_S \vec{J} \cdot d\vec{a}. \quad (1.11)$$

Make sure you know how to get this equation from (1.10)! While the probability current in more than one spatial dimension is also a vector, in our present one-dimensional case, it has just one component. The conservation equation is the analog of (1.10):

$$\frac{\partial J}{\partial x} + \frac{\partial \rho}{\partial t} = 0. \quad (1.12)$$

You can check that this equation holds using the above formula for $J(x, t)$, the formula for $\rho(x, t)$, and the Schrödinger equation. The integral version is formulated by first defining the probability $P_{ab}(t)$ of finding the particle in the interval $x \in [a, b]$

$$P_{ab}(t) \equiv \int_a^b dx |\Psi(x, t)|^2 = \int_a^b dx \rho(x, t). \quad (1.13)$$

You can then quickly show that

$$\frac{dP_{ab}}{dt}(t) = J(a, t) - J(b, t). \quad (1.14)$$

Here $J(a, t)$ denotes the rate at which probability flows in (in units of one over time) at the left boundary of the interval, while $J(b, t)$ denotes the rate at which probability flows out at the right boundary of the interval.

It is sometimes easier to work with wavefunctions that are not normalized. The normalization can be performed if needed. We will thus refer to wavefunctions in general without assuming normalization, otherwise we will call them *normalized* wavefunction. In this spirit, two wavefunctions Ψ_1 and Ψ_2 solving the Schrödinger equation are *declared* to be physically equivalent if they differ by multiplication by a complex *number*. Using the symbol \sim for equivalence, we write

$$\Psi_1 \sim \Psi_2 \iff \Psi_1(x, t) = \alpha \Psi_2(x, t), \quad \alpha \in \mathbb{C}. \quad (1.15)$$

If the wavefunctions Ψ_1 and Ψ_2 are normalized they are equivalent if they differ by an overall constant phase:

$$\text{Normalized wavefunctions: } \Psi_1 \sim \Psi_2 \iff \Psi_1(x, t) = e^{i\theta} \Psi_2(x, t), \quad \theta \in \mathbb{R}. \quad (1.16)$$

2 Stationary Solutions

In a large class of problems the Schrödinger potential $V(x, t)$ has no time dependence and it is simply a function $V(x)$ of position. We focus on that case now. The Schrödinger equation (1.2) can be written more briefly as

$$i\hbar \frac{\partial \Psi}{\partial t}(x, t) = \hat{H} \Psi(x, t), \quad (2.1)$$

where we have introduced the Hamiltonian operator \hat{H} :

$$\hat{H} \equiv -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x). \quad (2.2)$$

\hat{H} is an operator in the sense that it acts on functions of x and t to give functions of x and t : it acts on the space of complex functions, a space that contains wavefunctions. Note that $V(x)$ acts just by multiplication. Note that the operator \hat{H} is time independent – it does not involve time at all.

A **stationary** state of energy $E \in \mathbb{R}$ is a state $\Psi(x, t)$ that takes the form

$$\Psi(x, t) = e^{-iEt/\hbar} \psi(x), \quad (2.3)$$

where $\psi(x) \in \mathbb{C}$ is a function of x only that solves an equation that will be discussed below. All the time dependence of the stationary state is carried by the exponential prefactor. Such a state is called stationary because physical observables of the state are actually time independent. Consider, for example, the norm of the state. We see that the time dependence drops out

$$P(x, t) = \Psi^*(x, t) \Psi(x, t) = e^{+iEt/\hbar} \psi^*(x) e^{-iEt/\hbar} \psi(x) = \psi^*(x) \psi(x) = |\psi(x)|^2. \quad (2.4)$$

Had the energy E been a complex number $E = E_0 - i\Gamma$, with E_0 and Γ real, the time dependence would not drop out:

$$\begin{aligned} P(x, t) &= \Psi^*(x, t) \Psi(x, t) = e^{+iE^*t/\hbar} \psi^*(x) e^{-iEt/\hbar} \psi(x) \\ &= e^{i(E^*-E)t/\hbar} \psi^*(x) \psi(x) = e^{-2\Gamma t/\hbar} |\psi(x)|^2. \end{aligned} \quad (2.5)$$

This kind of state is not acceptable: the normalization cannot be preserved in time.

Let us consider the equation that $\psi(x)$ must satisfy. Plugging (2.3) into (2.1) we find

$$i\hbar \frac{\partial}{\partial t} e^{-iEt/\hbar} \psi(x) = \hat{H} e^{-iEt/\hbar} \psi(x), \quad (2.6)$$

The time derivative on the left-hand side only acts on the exponential and the \hat{H} operator on the right-hand side can be moved through the exponential (it commutes with it!) so we get

$$i\hbar \left(\frac{\partial}{\partial t} e^{-iEt/\hbar} \right) \psi(x) = e^{-iEt/\hbar} \hat{H} \psi(x). \quad (2.7)$$

Taking the derivative and canceling the exponentials we get

$$i\hbar \left(-i \frac{E}{\hbar} \right) \psi(x) = \hat{H} \psi(x), \quad (2.8)$$

which we write as

$$\boxed{\hat{H} \psi(x) = E \psi(x)}. \quad (2.9)$$

Recalling the expression for the Hamiltonian we have

$$\boxed{\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right) \psi(x) = E \psi(x)}. \quad (2.10)$$

Note that the derivatives along x need not be denoted as partial derivatives since the functions they act on have no other argument except x . Using primes to denote derivatives with respect to the argument, the above equation is

$$\boxed{-\frac{\hbar^2}{2m} \psi''(x) + V(x) \psi(x) = E \psi(x)}. \quad (2.11)$$

This is the equation for ψ that makes $\Psi(x, t) = e^{-iEt/\hbar} \psi(x)$ a stationary state of energy E . Any of the three boxed equations above is referred to as the **time-independent** Schrödinger equation.

Since the time-independent Schrödinger equation is a second-order differential equation in x , a solution is completely determined by the value of ψ and ψ' at some point x_0 ($\neq \infty$). If $V(x)$

is nowhere infinite, $\psi = \psi' = 0$ at some point implies $\psi = 0$ everywhere. Alternatively, if we know the solution for *any* size x -interval, the full solution is fully determined. A full solution means finding all the values E for which acceptable solutions $\psi(x)$ exist and, of course, finding those solutions for each E .

A solution $\psi(x)$ associated with an energy E is called an **energy eigenstate** of energy E . The set of all allowed values of E is called the **spectrum** of the Hamiltonian \hat{H} . A **degeneracy** in the spectrum occurs when there is more than one solution $\psi(x)$ for a given value of the energy.

The solutions depend on the properties of the potential $V(x)$. We will consider potentials $V(x)$ that can fail to be continuous (but are piece-wise continuous, like the finite square well) and can fail to be bounded (like the potential for the harmonic oscillator). We allow delta function contributions in the potential but do not allow worse singularities, such as squares or derivatives of delta functions. We allow hard walls, as in the infinite square-well.

On the wavefunction we impose the following regularity condition:

$$\psi(x) \text{ is continuous and bounded and its derivative } \psi'(x) \text{ is bounded.} \quad (2.12)$$

We do not impose the requirement that $\psi(x)$ be normalizable. This would be too restrictive. There are energy eigenstates that are not normalizable. Momentum eigenstates of a free particle are also not normalizable. Solutions for which ψ is not normalizable do not have a direct physical interpretation, but are very useful: suitable superpositions of them give normalizable solutions that can represent a particle.

In the spectrum of a Hamiltonian, localized energy eigenstates are particularly important. This motivates the definition:

$$\text{An energy eigenstate } \psi(x) \text{ is a } \mathbf{bound} \text{ state if } \psi(x) \rightarrow 0 \text{ when } |x| \rightarrow \infty. \quad (2.13)$$

Since a normalizable eigenstate must have a wavefunction that vanishes as $|x| \rightarrow \infty$, a bound state is just a normalizable eigenstate.

The eigenstates of \hat{H} provide a useful set of functions. Let us denote the possible energies by E_n with $n = 1, 2, \dots$, ordered as follows

$$E_1 \leq E_2 \leq E_3 \leq \dots \quad (2.14)$$

and let the corresponding eigenstates be $\psi_n(x)$, with

$$\hat{H}\psi_n(x) = E_n \psi_n(x), \quad n \geq 1. \quad (2.15)$$

For simplicity we discuss the case when the spectrum is denumerable so that, as above, we can label the states and energies with the integer label n . In general a potential $V(x)$ can result in a

spectrum that contains a discrete part and a continuous part. The discrete part is denumerable but the continuous part is not. The formulae we will write below require some modification when there spectrum contains a continuous part. The eigenstates of the continuum spectrum are not normalizable.

It is a known result about differential equations that for rather general potentials the \hat{H} eigenstates $\psi_n(x)$ can be chosen to be orthonormal. What does it mean for two functions to be orthogonal? Orthogonal vectors have a vanishing dot product, where the dot product is a (clever) rule to obtain a single number from two vectors. For two functions f_1 and f_2 an *inner* product can be defined by integrating the product function $f_1 f_2$ over all x , thus giving us a number. Since our functions are complex valued, a small modification is needed: the inner product of f_1 and f_2 is taken to be $\int f_1^* f_2$. The functions f_1 and f_2 are orthogonal if this integral vanishes. An orthonormal set of functions is one in which each function is orthogonal to all others, while its inner product with itself gives one (this requires the complex conjugation in the definition, can you see that?). As a result, orthonormality means that

$$\boxed{\text{Orthonormality: } \int_{-\infty}^{\infty} dx \psi_m^*(x) \psi_n(x) = \delta_{m,n}.} \quad (2.16)$$

Recall that the Kronecker delta $\delta_{m,n}$ is defined to be zero if $m \neq n$ and one otherwise.

The energy eigenstates are also *complete* in the sense that any reasonable (see (2.12) wavefunction $\psi(x)$ can be expanded as a superposition of energy eigenstates. Namely, there exist complex numbers b_n such that

$$\psi(x) = \sum_{n=1}^{\infty} b_n \psi_n(x), \quad b_n \in \mathbb{C}. \quad (2.17)$$

This is a very powerful statement: it means that if the energy eigenstates are known, the general solution of the Schrödinger equation is known. Indeed assume that the wavefunction at time equal zero is the $\psi(x)$ above. Then we have

$$\boxed{\Psi(x, t=0) = \psi(x) = \sum_{n=1}^{\infty} b_n \psi_n(x).} \quad (2.18)$$

If this wavefunction is normalized then we have

$$\int_{-\infty}^{\infty} dx \psi^*(x) \psi(x) = 1 \rightarrow \sum_{n=1}^{\infty} |b_n|^2 = 1. \quad (2.19)$$

We now claim that the wavefunction at all times can be written down immediately

$$\boxed{\Psi(x, t) = \sum_{n=1}^{\infty} b_n e^{-iE_n t/\hbar} \psi_n(x).} \quad (2.20)$$

To prove that this is the solution we first note that we have produced a solution to the Schrödinger equation: this follows by linearity because each term in the above sum is a solution (a stationary state). Second, the solution reduces for $t = 0$ to the correct value $\Psi(x, t = 0)$ in (2.18). By the first remark below (1.2) this is all that is needed.

It should be emphasized that the superposition of stationary states is generally *not* a stationary state. The expansion coefficients b_n used above can be calculated explicitly if we know the energy eigenstates. Indeed using (2.16) and (2.17) a one-line computation (do it!) gives

$$b_n = \int_{-\infty}^{\infty} dx \psi_n^*(x) \psi(x). \quad (2.21)$$

A curious identity can be derived by substituting this result back into (2.17):

$$\psi(x) = \sum_{n=1}^{\infty} \left(\int_{-\infty}^{\infty} dx' \psi_n^*(x') \psi(x') \right) \psi_n(x) = \int_{-\infty}^{\infty} dx' \left(\sum_{n=1}^{\infty} \psi_n^*(x') \psi_n(x) \right) \psi(x'), \quad (2.22)$$

where we interchanged the order of integration and summation (a safe operation in most cases!). The above equation is of the form

$$f(x) = \int_{-\infty}^{\infty} dx' K(x', x) f(x') \quad (2.23)$$

and is supposed to hold for any function $f(x)$. It is intuitively clear that that $K(x', x)$ must vanish for $x' \neq x$ for otherwise we could cook up a contradiction by choosing a peculiar function $f(x)$. Taking $f(x) = \delta(x - x_0)$ the equation gives

$$\delta(x - x_0) = \int dx' K(x', x) \delta(x' - x_0) = K(x_0, x). \quad (2.24)$$

We therefore conclude that $K(x', x) = \delta(x - x')$ (recall that $\delta(x) = \delta(-x)$). Back in (2.22) we thus find

$$\text{Completeness: } \sum_{n=1}^{\infty} \psi_n^*(x') \psi_n(x) = \delta(x - x'). \quad (2.25)$$

Let us compare the completeness relation above with the orthonormality relation (2.16). In the completeness relation we set equal the labels of the eigenfunctions and sum over them while keeping the two position arguments fixed. In the orthogonality relation we set equal the position arguments of the eigenfunctions and integrate (sum) over them while keeping the two labels fixed. On the right-hand sides we find “delta functions”: a Kronecker delta setting equal the two labels in the orthonormality relation and a true delta function setting equal the two positions in the completeness relation. The two relations are obtained from each other by exchange of labels: position labels and energy labels. This is a neat duality!

It is fun to calculate the expectation value of the Hamiltonian in the solution $\Psi(x, t)$ in (2.20). For arbitrary time-independent operators \hat{A} one defines the (generally) time-dependent expectation value on a *normalized* state Ψ by

$$\langle \hat{A} \rangle_{\Psi}(t) \equiv \int_{-\infty}^{\infty} dx \Psi^*(x, t)(\hat{A}\Psi(x, t)) . \quad (2.26)$$

What happens when we take the operator to be \hat{H} ? Using (2.20) twice, we get

$$\begin{aligned} \langle \hat{H} \rangle_{\Psi}(t) &= \int_{-\infty}^{\infty} dx \Psi^*(x, t)(\hat{H}\Psi(x, t)) \\ &= \sum_{n,n'} \int_{-\infty}^{\infty} dx b_n^* e^{iE_n t/\hbar} \psi_n^*(x) b_{n'} e^{-iE'_{n'} t/\hbar} \hat{H} \psi_{n'}(x) \\ &= \sum_{n,n'} b_n^* b_{n'} E_{n'} e^{i(E_n - E_{n'})t/\hbar} \int_{-\infty}^{\infty} dx \psi_n^*(x) \psi_{n'}(x) \\ &= \sum_{n,n'} b_n^* b_{n'} E_{n'} e^{i(E_n - E_{n'})t/\hbar} \delta_{n,n'} , \end{aligned} \quad (2.27)$$

so that we get

$$\langle \hat{H} \rangle_{\Psi}(t) = \sum_{n=1}^{\infty} |b_n|^2 E_n . \quad (2.28)$$

The expectation value of the Hamiltonian is time-independent: this is the quantum version of energy conservation. This is the expected value of the energy: a weighted sum of the possible energies with weights the norm-squared of the expansion coefficients.

If the wavefunction $\Psi(x, t)$ is not normalized but is normalizable, then the wavefunction

$$\frac{\Psi(x, t)}{\sqrt{\int dx \Psi^* \Psi}} \quad (2.29)$$

is normalized. We can thus use this normalized wavefunction in the definition on $\langle \hat{A} \rangle$ to find the expectation value is given by

$$\langle \hat{A} \rangle_{\Psi}(t) \equiv \frac{\int_{-\infty}^{\infty} dx \Psi^*(x, t)(\hat{A}\Psi(x, t))}{\int dx \Psi^*(x, t)\Psi(x, t)} . \quad (2.30)$$

This formula can be used for any normalizable Ψ . If the Ψ is normalized the formula reduces to the earlier expression for $\langle \hat{A} \rangle$.

Another operator often used to explore the physics of states is the momentum operator \hat{p} . Acting on wavefunctions that depend on a coordinate x it takes the form of a differential operator:

$$\hat{p} \equiv \frac{\hbar}{i} \frac{\partial}{\partial x} . \quad (2.31)$$

3 Properties of energy eigenstates in one dimension

In order to simplify our notation we rewrite the time-independent Schrödinger equation (2.10) as follows

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - V(x))\psi = 0. \quad (3.1)$$

We then define energy-like quantities \mathcal{E} and \mathcal{V} using a common rescaling factor:

$$\mathcal{E} \equiv \frac{2m}{\hbar^2}E, \quad \mathcal{V}(x) \equiv \frac{2m}{\hbar^2}V(x). \quad (3.2)$$

With this the Schrödinger equation (3.1) becomes

$$\psi'' + (\mathcal{E} - \mathcal{V}(x))\psi = 0. \quad (3.3)$$

We are now ready to consider a basic result: in a one-dimensional potential there cannot be two or more bound states for any given energy.

Theorem 1. There is no degeneracy for bound states in one-dimensional potentials.

Proof. Suppose there is such degeneracy so that there are $\psi_1(x)$ and $\psi_2(x)$, different from each other and both corresponding to the same energy E , thus same value of \mathcal{E} . If so, we have that the following equations hold

$$\begin{aligned} \psi_1'' + (\mathcal{E} - \mathcal{V}(x))\psi_1 &= 0, \\ \psi_2'' + (\mathcal{E} - \mathcal{V}(x))\psi_2 &= 0. \end{aligned} \quad (3.4)$$

Multiplying the top equation by ψ_2 and the bottom one by ψ_1 and subtracting them we find

$$\psi_2\psi_1'' - \psi_1\psi_2'' = 0. \quad (3.5)$$

The left-hand side is actually a derivative

$$(\psi_2\psi_1' - \psi_1\psi_2')' = 0. \quad (3.6)$$

It follows from this that the expression inside the parenthesis must be a constant c ,

$$\psi_2\psi_1' - \psi_1\psi_2' = c. \quad (3.7)$$

The constant can be evaluated by examining the left-hand side for $|x| \rightarrow \infty$. We then have that $\psi_1 \rightarrow 0$ and $\psi_2 \rightarrow 0$, since they are bound states, while the derivatives are bounded, as assumed in (2.12). It follows that the left-hand side vanishes as $|x| \rightarrow \infty$ and therefore $c = 0$. We thus have

$$\psi_2\psi_1' = \psi_1\psi_2' \quad \rightarrow \quad \frac{\psi_1'}{\psi_1} = \frac{\psi_2'}{\psi_2} \quad \rightarrow \quad \frac{d}{dx}(\ln \psi_1 - \ln \psi_2) = 0. \quad (3.8)$$

This implies that we have for some constant c'

$$\ln \psi_1 = \ln \psi_2 + \ln c' \quad \rightarrow \quad \psi_1(x) = c' \psi_2(x). \quad (3.9)$$

We have thus shown that the wavefunctions ψ_1 and ψ_2 are equivalent. In contradiction with the initial assumption, they are the same energy eigenstate. This concludes the proof.

For our second theorem we show that the reality of \mathcal{V} allows us to work with real wavefunctions.

Theorem 2. The energy eigenstates $\psi(x)$ can be chosen to be real.

Proof. Consider our main equation and a possibly complex wavefunction that correspond

$$\psi'' + (\mathcal{E} - \mathcal{V}(x))\psi = 0, \quad (3.10)$$

Since $(\psi'')^* = (\psi^*)''$ the complex conjugation of the above equation gives

$$(\psi^*)'' + (\mathcal{E} - \mathcal{V}(x))\psi^* = 0. \quad (3.11)$$

So ψ^* if different from ψ defines a degenerate solution. By superposition we can then get two real (degenerate) solutions

$$\psi_r \equiv \frac{1}{2}(\psi + \psi^*), \quad \psi_{im} \equiv \frac{1}{2i}(\psi - \psi^*). \quad (3.12)$$

These are, of course, the real and imaginary parts of ψ .

If we are dealing with bound states of one-dimensional potentials more can be said: any such solution is, up to a phase, *equal* to a real solution. Indeed, the absence of degenerate bound states means that the two real solutions ψ_r and ψ_{im} must be equal up to a constant that can only be real:

$$\psi_{im} = c \psi_r, \quad \text{with } c \in \mathbb{R} \quad (3.13)$$

It then follows that $\psi = \psi_r + i\psi_{im} = (1 + ic)\psi_r$. Writing $1 + ic = \sqrt{1 + c^2} e^{i\beta}$ with real beta, shows that ψ is, up to a phase β , equal to a real solution.

Theorem 3. If the potential is an even function of x : $V(-x) = V(x)$ the eigenstates can be chosen to be even or odd under $x \rightarrow -x$.

Proof. Again, we begin with our main equation

$$\psi''(x) + (\mathcal{E} - \mathcal{V}(x))\psi(x) = 0. \quad (3.14)$$

Recall that primes denote here derivative with respect to the argument, so $\psi''(x)$ means the function “second-derivative-of- ψ ” evaluated at x . Similarly $\psi''(-x)$ means the function “second-derivative-of- ψ ” evaluated at $-x$. Thus we can change x for $-x$ with impunity in the above equation getting

$$\psi''(-x) + (\mathcal{E} - \mathcal{V}(x))\psi(-x) = 0, \quad (3.15)$$

where we used that V , and thus \mathcal{V} , is even. We now want to make clear that the above equation implies that $\psi(-x)$ is another solution of the Schrödinger equation with the same energy. For this let us define

$$\varphi(x) \equiv \psi(-x) \rightarrow \frac{d}{dx}\varphi(x) = \psi'(-x) \cdot (-1). \quad (3.16)$$

Taking a second derivative and using (3.15)

$$\frac{d^2}{dx^2}\varphi(x) = \psi''(-x) = -(\mathcal{E} - \mathcal{V}(x))\varphi(x), \quad (3.17)$$

so that indeed $\varphi(x) = \psi(-x)$ provides a degenerate solution to the Schrödinger equation:

$$\frac{d^2}{dx^2}\varphi(x) + (\mathcal{E} - \mathcal{V}(x))\varphi(x) = 0. \quad (3.18)$$

Equipped with the degenerate solutions $\psi(x)$ and $\psi(-x)$ we can now form symmetric (s) and antisymmetric (a) combinations that are, respectively, even and odd under $x \rightarrow -x$:

$$\psi_s(x) \equiv \frac{1}{2}(\psi(x) + \psi(-x)), \quad \psi_a(x) \equiv \frac{1}{2}(\psi(x) - \psi(-x)). \quad (3.19)$$

These are the solutions claimed to exist in Theorem 3.

Again, if we focus on bound states of one-dimensional potentials the absence of degeneracy implies that $\psi(x)$ and $\psi(-x)$ must be the same solution. Because of Theorem 2 we can choose $\psi(x)$ to be real and thus we must have

$$\psi(-x) = c\psi(x), \quad \text{with } c \in \mathbb{R}. \quad (3.20)$$

Letting $x \rightarrow -x$ in the above equation we get $\psi(x) = c\psi(-x) = c^2\psi(x)$ from which we learn that $c^2 = 1$. The only possibilities are $c = \pm 1$. So $\psi(x)$ is *automatically* even or odd under $x \rightarrow -x$. Any one-dimensional bound state solution with an even potential *must be* either even or odd under $x \rightarrow -x$.

4 The nature of the spectrum

Consider the time-independent Schrödinger equation written as

$$\psi'' = -\frac{2m}{\hbar^2}(E - V(x))\psi. \quad (4.1)$$

We always have that $\psi(x)$ is continuous, otherwise ψ'' has singularities worse than delta functions and we would require potentials $V(x)$ that are worse than delta functions – something we will not consider. Consider now three possibilities concerning the potential:

1. $V(x)$ is continuous. In this case the continuity of $\psi(x)$ and (4.1) imply ψ'' is also continuous. This requires ψ' continuous.

2. $V(x)$ has finite jumps. In this case ψ'' has finite jumps (it is the multiplication of ψ with no jumps times V with jumps). But then ψ' can have no jumps (it is continuous, with non-continuous derivative).
3. $V(x)$ contains delta functions. In this case ψ'' also contains delta functions (it is the multiplication of the continuous ψ times a delta function in V). Thus ψ' has finite jumps.
4. $V(x)$ contains a hard wall. A potential that is finite immediately to the left of $x = a$ and becomes infinite for $x > a$ is said to have a hard wall at $x = a$. In such a case, the wavefunction will vanish for $x \geq a$. The slope ψ' will be finite as $x \rightarrow a$ from the left, and will vanish for $x > a$. Thus ψ' is discontinuous at the wall.

In conclusion

Both ψ and ψ' are continuous unless the potential has delta functions or hard walls in which cases ψ' may have finite jumps. (4.2)

The origin of the discrete and continuous spectrum can be seen from simple examples. We have three situations to discuss, as shown in Figure 1 as (a), (b), and (c). We will consider the number of parameters needed to write a solution and the number of constraints due to boundary conditions. Without loss of generality we can consider real solutions, and therefore the parameters will be real.

- (a) Here the energy E is below the potential far to the left and far to the right, but not in the middle. On the left the solution must be a decaying exponential $\alpha_1 \exp(-\kappa|x|)$, where α_1 is a constant to be determined and κ is known if the energy E is known. So thus far we got one unknown constant α_1 . In the middle region where $E > V$ the solution is oscillatory $\alpha_2 \cos kx + \alpha_3 \sin kx$, with two unknown constants α_2 and α_3 , and k determined if E is known. Finally to the right we have a solution $\alpha_4 \exp(-\kappa x)$ since the wavefunction must vanish as $x \rightarrow \infty$. So we got four (real) unknown constants α_i , $i = 1, 2, 3, 4$. Since ψ and $c\psi$ are the same solution we can scale the solution and thus we only have three unknown constants to determine. There are, however, four constraints from boundary conditions: the continuity of ψ and ψ' at each of the two interfaces. With three coefficients and four conditions we cannot expect a solution to exist. If we view the energy E , however, as unknown, then we have four unknowns and four conditions. Indeed solutions exist for discrete values of the energy. We get a discrete spectrum.
- (b) Here we have one unknown constant for the solution to the left of the interface (multiplying a decaying exponential) and two unknown constants for the oscillatory solution to the right of the interface, for a total of three unknowns, or just two unknowns once the overall scale

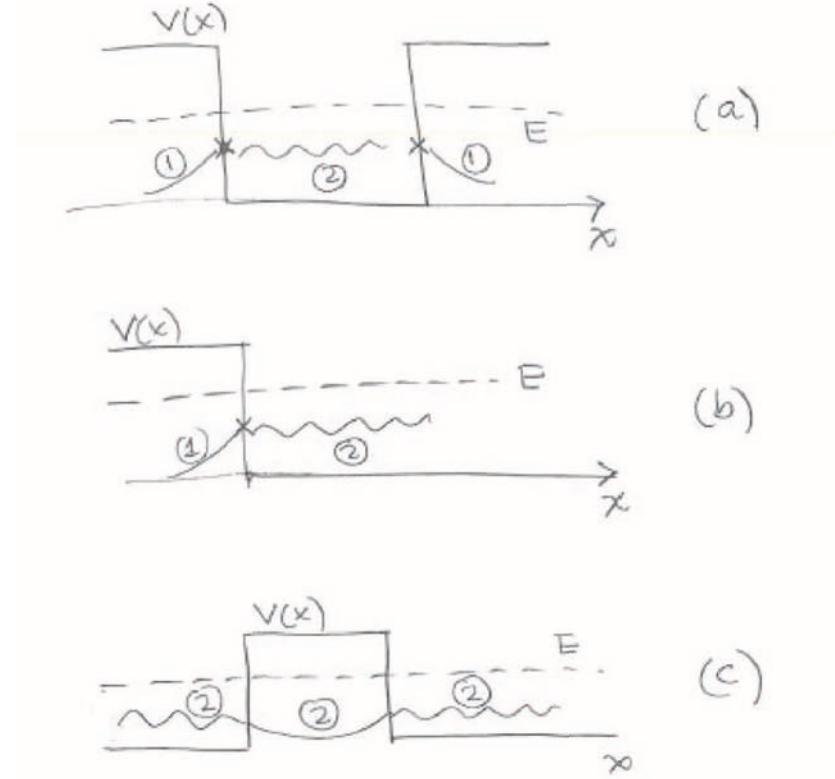


Figure 1: Discussing the number of constants needed to specify a solution. (a) Energy is smaller than the potential for $x \rightarrow \pm\infty$. (b) Energy is smaller than the potential for $x \rightarrow -\infty$ and larger than the potential for $x \rightarrow \infty$. (c) Energy is larger than the potential for $x \rightarrow \pm\infty$.

freedom is accounted in. We also have two boundary conditions at the interface. So we can expect a solution. Indeed there should be a solution for each value of the energy. The spectrum here is continuous and non-degenerate.

- (c) Two constants are needed here in each of the three regions: they multiply sines and cosines to the left and to the right, and multiply the two exponentials in the middle. Thus six constants and due to scaling just five unknowns. We still have four boundary conditions so there should be solutions. In fact, there are two solutions for each energy. We can understand this as follows. Think of using just one coefficient to the far left, say the coefficient multiplying the sine function. With one less coefficient we have the same number of unknowns as constraints so we should get one solution (for any E). We get another solution if we use the cosine function to the far left. So we have two solutions for each energy. The spectrum is continuous and doubly degenerate.

Figure 2 illustrates the spectrum of the Hamiltonian for a rather generic type of potential. Here V_+ is the top asymptote of the potential, V_- is the bottom asymptote of the potential, and V_0 is the lowest value of the potential. In the figure we indicate the type of spectrum for energies in the various intervals defined: $E > V_+$, then $V_- < E < V_+$, then $V_0 < E < V_-$ and finally $E < V_0$.

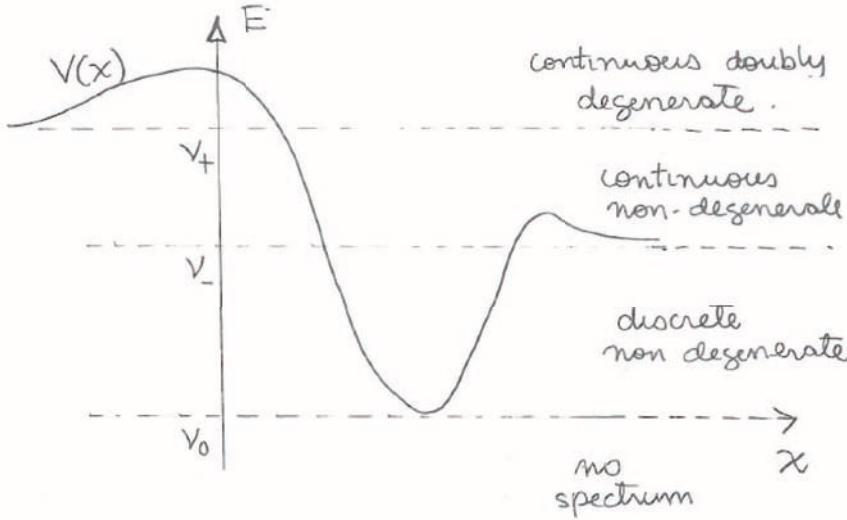


Figure 2: A generic potential and the type of spectrum for various energy ranges.

A **node** in a wavefunction is a point x_0 where $\psi(x_0) = 0$ (a zero of ψ) and $\psi'(x_0) \neq 0$. For a bound state the zeroes are nodes or the points at infinity (where typically $\psi' \rightarrow 0$).

Theorem 4 For the discrete bound-state spectrum of a one-dimensional potential let the allowed energies be $E_1 < E_2 < E_3 < \dots$ with E_1 the ground state energy. Let the associated energy eigenstates be $\psi_1, \psi_2, \psi_3, \dots$. The wavefunction ψ_1 has no nodes, ψ_2 has one node, and each consecutive wavefunction has one additional node. In conclusion ψ_n has $n - 1$ nodes.

We will not prove this theorem here. In fact you will show in the homework that ψ_{k+1} has at least one node between two consecutive zeroes of ψ_k . This implies that ψ_{k+1} has at least one more node than ψ_k . This can be illustrated in Figure 3 that shows a bound state $\psi_4(x)$ with three nodes at x_1, x_2 , and x_3 and zeroes at $x = -\infty$ and $x = \infty$. For ψ_5 there must be a node w_1 in $(-\infty, x_1]$, a node $w_2 \in (x_1, x_2)$ and so on until a last node $w_4 \in (x_3, \infty)$.

Example: Potential with five delta functions. We will discuss the bound states of the Schrödinger equation with potential

$$V(x) = -V_0 a \sum_{n=-2}^2 \delta(x - na). \quad (4.3)$$

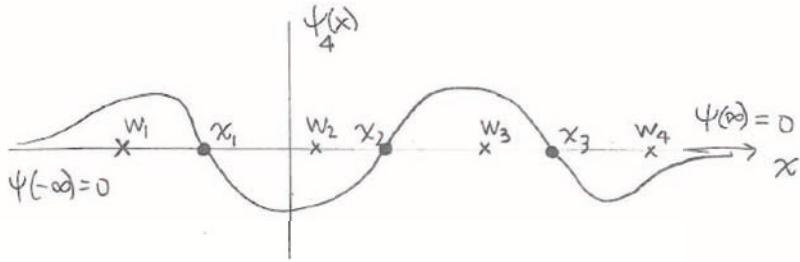


Figure 3: A wavefunction ψ_4 with three nodes (x_1, x_2, x_3) and zeroes at $x \pm \infty$. The next wavefunction ψ_5 must have four nodes, with positions indicated by w_1, w_2, w_3 and w_4 .

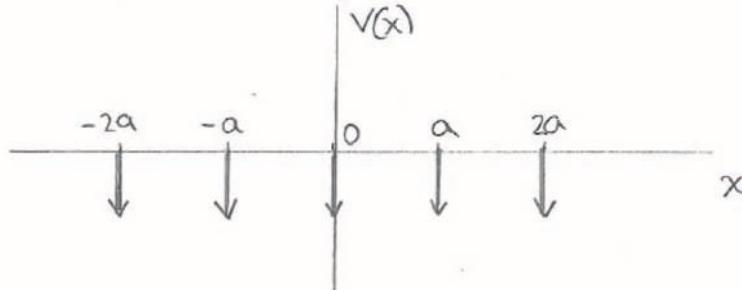


Figure 4: A potential $V(x)$ with five downwards pointing delta-functions.

This potential has delta functions at x equal to $-2a, -a, 0, a$, and $2a$, as shown in Figure 4.

We first examine the effect of the delta functions on the eigenstates. We will see that they produce discontinuities in ψ' at the position of the delta functions. We begin with the Schrödinger equation

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + V(x)\psi(x) = E\psi(x), \quad (4.4)$$

and integrate this equation from $a - \epsilon$ to $a + \epsilon$, where ϵ is a small value that we will take down to zero. By doing this we will get one out of the five delta functions to fire. We find

$$-\frac{\hbar^2}{2m} \int_{a-\epsilon}^{a+\epsilon} dx \frac{d^2\psi}{dx^2} + \int_{a-\epsilon}^{a+\epsilon} dx V(x)\psi(x) = E \int_{a-\epsilon}^{a+\epsilon} dx \psi(x). \quad (4.5)$$

The first term involves a total derivative, the second term just picks up the delta function at $x = a$, and the right hand side is evaluated by noting that since ψ is continuous its value at $x = a$ gives the leading contribution:

$$-\frac{\hbar^2}{2m} \frac{d\psi}{dx} \Big|_{a-\epsilon}^{a+\epsilon} - V_0 a \int_{a-\epsilon}^{a+\epsilon} dx \delta(x-a)\psi(x) = E(2\epsilon)\psi(a) + \mathcal{O}(\epsilon^2). \quad (4.6)$$

In the limit as $\epsilon \rightarrow 0$ we will denote $a + \epsilon$ as a^+ and $a - \epsilon$ as a^- . These labels are needed since ψ' has to be discontinuous at x . Indeed, we get

$$-\frac{\hbar^2}{2m} (\psi'(a^+) - \psi'(a^-)) - V_0 a \psi(a) = 0. \quad (4.7)$$

This implies that the discontinuity $\Delta\psi'$ of ψ' is given by

$$\Delta\psi'(a) \equiv \psi'(a^+) - \psi'(a^-) = \frac{2m}{\hbar^2} (-V_0 a) \psi(a).$$

The discontinuity of ψ' at the position of the delta function is proportional to the value of ψ at this point. The constant of proportionality is linear on the strength $V_0 a$ of the delta function. It follows that if the delta function of the potential is at a point where ψ vanishes then both ψ and ψ' are continuous and the delta function has no effect.

Let us now focus on bound states. These will be states with $E < 0$. The Schrödinger equation away from the delta functions is just

$$\psi'' = -\frac{2mE}{\hbar^2} \psi = \kappa^2 \psi, \text{ with } \kappa^2 \equiv -\frac{2mE}{\hbar^2} > 0, \kappa > 0. \quad (4.9)$$

The solutions are therefore the linear combinations

$$\psi(x) = ae^{-\kappa x} + be^{\kappa x}, \quad (4.10)$$

with a and b real constants to be determined (recall the wavefunction can be taken to be real). In Figure 5 we show these functions for $a > b > 0$. Note the the curves intersect just once. It follows that the wavefunction will never have a zero if a and b have the same sign and it will have exactly one zero if a and b have opposite signs.

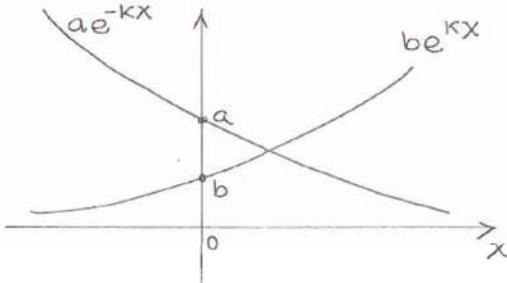


Figure 5: Plots of $ae^{-\kappa x}$ and $be^{\kappa x}$ with $a, b > 0$. This can be used to show that any linear superposition of these two functions can at most have one zero.

Let us then make the following remarks:

1. There cannot be zeroes of the wavefunction for $x \geq 2a$ (nor for $x \leq -2a$). For $x \geq 2a$ the solution, if non vanishing, must be of the form $ce^{-\kappa x}$. This can only have a zero if $c = 0$. In this case the wavefunction would vanish identically for $x \geq 2a$. This does not look good. Since $\psi(2a) = 0$ then ψ' is not discontinuous and, by continuity, a bit to the left of $2a$ both ψ and ψ' vanish. This is enough to make the solution vanish over the next interval $x \in (a, 2a)$. Continuing in this way we find that the solution for ψ would have to be zero everywhere. This is not acceptable.
2. There is at most one zero in between each pair of contiguous δ -functions. This follows because the solution must take the form (4.10) and we argued that such function can at most have one zero.
3. Zeroes appear at $x = 0$ for all the antisymmetric bound states. In those cases, there cannot be another zero in the interval $[-a, a]$. Zeroes may appear at $x = \pm a$, but this is presumably not generic. There are at most five bound states because the maximum number of nodes is four; one in between each delta function. All these five bound states exist if the delta functions are strong enough. The ground state is even, has no nodes and presumably looks like the one drawn in Figure 6.

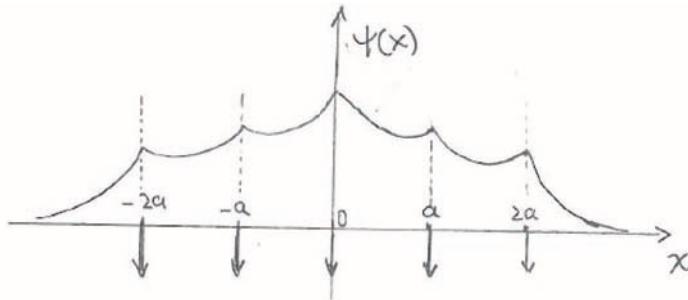


Figure 6: A sketch of the ground state wavefunction.

Exercise. Sketch the expected shapes of the four excited bound states of the potential.

5 Variational Principle

Consider a system with Hamiltonian \hat{H} and focus on the time-independent Schrödinger equation:

$$\hat{H}\psi(\vec{x}) = E\psi(\vec{x}). \quad (5.11)$$

Let us assume that the system is such that it has a collection of energy eigenstates that are normalizable. This collection includes a ground state with ground state energy E_{gs} . Note the

use of \vec{x} : our discussion applies to quantum systems in any number of spatial dimensions. Our first goal is to learn something about the ground state energy without solving the Schrödinger equation nor trying to figure out the ground state wavefunction.

For this purpose, consider an *arbitrary* normalized wavefunction $\psi(\vec{x})$:

$$\int d\vec{x} \psi^*(\vec{x}) \psi(\vec{x}) = 1. \quad (5.12)$$

By arbitrary we mean a wavefunction that need not satisfy the time-independent Schrödinger equation, a wavefunction that need not be an energy eigenstate. Then we claim the ground state energy E_{gs} of the Hamiltonian is smaller or equal than the expectation value of \hat{H} in this arbitrary normalized ψ , namely,

$$E_{gs} \leq \langle \hat{H} \rangle_\psi = \int d\vec{x} \psi^*(\vec{x}) \hat{H} \psi(\vec{x}), \quad \text{Normalized } \psi. \quad (5.13)$$

The wavefunction $\psi(\vec{x})$ here is sometimes called a *trial* wavefunction. When the right-hand side of the above inequality is evaluated we get an energy and learn that the ground state energy must be smaller or equal to the value we get. Thus any trial wavefunction provides an *upper bound* for the ground state energy. Better and better trial wavefunctions will produce lower and lower upper bounds. Note that if the trial wavefunction was set equal to the (unknown) ground-state wavefunction, the expectation value of \hat{H} becomes exactly E_{gs} and the inequality is saturated.

Let us prove (5.13). For simplicity, we will consider here the case where the energy eigenstates $\psi_n(\vec{x})$ of \hat{H} are denumerable and their corresponding energies E_n are ordered as

$$E_{gs} = E_1 \leq E_2 \leq E_3 \leq \dots. \quad (5.14)$$

Of course $\hat{H}\psi_n = E_n\psi_n$. Since the energy eigenstates are complete, any trial wavefunction can be expanded in terms of them (see (2.18)):

$$\psi(\vec{x}) = \sum_{n=1}^{\infty} b_n \psi_n(\vec{x}). \quad (5.15)$$

Such a ψ is not an energy eigenstate in general. The normalization condition (5.12) gives us,

$$\sum_{n=1}^{\infty} |b_n|^2 = 1. \quad (5.16)$$

The evaluation of the right-hand side in (5.13) was done before in (2.28) so we have

$$\int d\vec{x} \psi^*(\vec{x}) \hat{H} \psi(\vec{x}) = \sum_{n=1}^{\infty} |b_n|^2 E_n. \quad (5.17)$$

Since $E_n \geq E_1$ for all n , we can replace the E_n on the above right-hand side for E_1 getting a smaller or equal value:

$$\int d\vec{x} \psi^*(\vec{x}) \hat{H} \psi(\vec{x}) \geq \sum_{n=1}^{\infty} |b_n|^2 E_1 = E_1 \sum_{n=1}^{\infty} |b_n|^2 = E_1 = E_{gs}, \quad (5.18)$$

where we used (5.16). This is in fact the claim in (5.13).

It is sometimes more convenient not to worry about the normalization of the trial wavefunctions. Given a trial wavefunction ψ that is not normalized, the wavefunction

$$\frac{\psi(x)}{\sqrt{N}} \quad \text{with } N = \int d\vec{x} \psi^*(\vec{x}) \psi(\vec{x}), \quad (5.19)$$

is normalized and can be used in (5.13). We therefore find that

$$E_{gs} \leq \frac{\int d\vec{x} \psi^*(\vec{x}) \hat{H} \psi(\vec{x})}{\int d\vec{x} \psi^*(\vec{x}) \psi(\vec{x})} \equiv \mathcal{F}[\psi]. \quad (5.20)$$

This formula can be used for trial wavefunctions that are not normalized. We also introduced the definition of the functional $\mathcal{F}[\psi]$. A functional is a machine that given a function gives us a number. Our result states that the ground state energy arises as the minimum value that the functional can take.

One application of this variational principle is to find good upper bounds for the ground state energy of quantum systems that are not exactly solvable. For this purpose it is useful to construct trial wavefunctions $\psi(\vec{x}; \beta_1, \beta_2, \dots, \beta_m)$ that depend on a set of parameters β . One then computes the expectation value $\langle \hat{H} \rangle_\psi$ which, of course, is a function of the parameters. Any random values for the parameters will give an upper bound for the ground state energy, but by minimizing $\langle \hat{H} \rangle_\psi$ over the parameter space we get the lowest possible upper bound consistent with the chosen form for the trial wavefunction.

Example. (Griffiths). Consider a one-dimensional problem with the delta function potential

$$V(x) = -\alpha \delta(x), \quad \alpha > 0. \quad (5.21)$$

In this problem the ground state energy is calculable exactly and one has

$$E_{gs} = -\frac{m\alpha^2}{2\hbar^2}. \quad (5.22)$$

So this problem is just for illustration. Consider an unnormalized gaussian trial wavefunction, with a real parameter β :

$$\psi(x) = e^{-\frac{1}{2}\beta^2 x^2}, \quad \int_{-\infty}^{\infty} dx \psi^2 = \frac{\sqrt{\pi}}{\beta}. \quad (5.23)$$

The functional \mathcal{F} in (5.20) is then¹

$$\begin{aligned}
\frac{\int dx \psi^*(x) \hat{H} \psi(x)}{\int dx \psi^*(x) \psi(x)} &= \frac{\beta}{\sqrt{\pi}} \int dx e^{-\frac{1}{2}\beta^2 x^2} \left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} - \alpha \delta(x) \right) e^{-\frac{1}{2}\beta^2 x^2} \\
&= \frac{\beta}{\sqrt{\pi}} \frac{\hbar^2}{2m} \int dx \left[\frac{d}{dx} e^{-\frac{1}{2}\beta^2 x^2} \right]^2 - \frac{\beta}{\sqrt{\pi}} \alpha \\
&= \frac{\beta}{\sqrt{\pi}} \frac{\hbar^2}{2m} \frac{\beta \sqrt{\pi}}{2} - \frac{\beta}{\sqrt{\pi}} \alpha \\
&= \frac{\beta^2 \hbar^2}{4m} - \frac{\beta}{\sqrt{\pi}} \alpha.
\end{aligned} \tag{5.24}$$

The first term on the last right-hand side is the kinetic energy and the second term is the potential energy. For any value of β the final expression above provides an upper bound for the ground state energy, and the best upper bound is the lowest one. We thus have that the ground state energy satisfies

$$E_{gs} \leq \text{Min}_\beta \left(\frac{\beta^2 \hbar^2}{4m} - \frac{\beta}{\sqrt{\pi}} \alpha \right). \tag{5.25}$$

The minimum is easily found

$$\beta = \frac{2m\alpha}{\hbar^2 \sqrt{\pi}} \quad \rightarrow \quad E_{gs} \leq -\frac{m\alpha^2}{\pi \hbar^2} = \frac{2}{\pi} \left(-\frac{m\alpha^2}{2\hbar^2} \right). \tag{5.26}$$

Comparing with (5.22) we see that the bound we found is in fact $\frac{2}{\pi} E_{gs} \simeq 0.64 E_{gs}$. The trial wavefunction brought us to about 64% of the correct value.

In the exercises you will develop the following results:

1. With trial wavefunctions orthogonal to the ground state, the functional \mathcal{F} gives upper bounds for the energy of the first excited state.
2. For any attractive one-dimensional potential (a nowhere positive potential that approaches zero at infinity) there is a bound state, namely, a state with energy less than zero.
3. We have shown that the functional $\mathcal{F}[\psi]$ has a minimum for ψ equal to the ground state wavefunction. Interestingly, this functional is stationary at each and every energy eigenstate. For eigenstates of energies higher than the ground state \mathcal{F} has a saddle point.

¹We use the integrals $\int du e^{-u^2} = \sqrt{\pi}$ and $\int du u^2 e^{-u^2} = \frac{1}{2}\sqrt{\pi}$.

6 Position and momentum

In quantum mechanics the position operator \hat{x} and the momentum operator \hat{p} do not commute. They satisfy the commutation relation

$$[\hat{x}, \hat{p}] = i\hbar. \quad (6.27)$$

When we deal with wavefunctions $\psi(x)$ the position operator acts on them in a simple way.² We define

$$\hat{x}\psi(x) \equiv x\psi(x). \quad (6.28)$$

In words the position operator acting on an x -dependent wavefunction simply multiplies the wavefunction by x . In quantum mechanics it is useful think of states as vectors and operators as matrices. A wavefunction for a particle on the box $0 \leq x \leq a$, for example can be thought as vector with many components, each one giving the value of the function at a specific point. To make this concrete one discretizes the space into small intervals of size ϵ such that $N\epsilon = a$. In that case we can represent the information in $\psi(x)$ in a large column vector

$$\psi(x) \longleftrightarrow \begin{pmatrix} \psi(0) \\ \psi(\epsilon) \\ \psi(2\epsilon) \\ \vdots \\ \psi(N\epsilon) \end{pmatrix}. \quad (6.29)$$

The $N + 1$ component column vector summarizes the values of the wavefunction at equally separated points. N is some kind of regulator: a precise description requires $N \rightarrow \infty$ or $\epsilon \rightarrow 0$. Associated with the description (6.29) the operator \hat{x} can be viewed as the $(N + 1) \times (N + 1)$ diagonal matrix

$$\hat{x} \longleftrightarrow \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \epsilon & 0 & \dots & 0 \\ 0 & 0 & 2\epsilon & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & N\epsilon \end{pmatrix}. \quad (6.30)$$

You can see that the action of the matrix \hat{x} on the vector (6.29) gives the vector

$$\begin{pmatrix} 0 \cdot \psi(0) \\ \epsilon \cdot \psi(\epsilon) \\ 2\epsilon \cdot \psi(2\epsilon) \\ \vdots \\ N\epsilon \cdot \psi(N\epsilon) \end{pmatrix}, \quad (6.31)$$

²The time dependence is irrelevant to the present discussion, which applies without changes to time-dependent wavefunctions $\Psi(x, t)$.

which is indeed the representation of $x\psi(x)$. Given our definition of the action of \hat{x} , expectation values in normalized states are naturally defined by

$$\langle \hat{x} \rangle \equiv \int dx' \psi^*(x') (x' \psi(x')) . \quad (6.32)$$

Are there eigenstates of the \hat{x} operator? Yes, but their are not normalizable. An eigenstate of \hat{x} must be a localized state, and the obvious candidate is a delta function. Defining

$$\psi_{x_0}(x) \equiv \delta(x - x_0) , \quad (6.33)$$

we verify that

$$\hat{x}\psi_{x_0}(x) = x\psi_{x_0}(x) = x\delta(x - x_0) = x_0\delta(x - x_0) = x_0\psi_{x_0}(x) , \quad (6.34)$$

confirming that $\psi_{x_0}(x)$ is an eigenstate of \hat{x} with eigenvalue x_0 . A delta function cannot be normalized, so the position eigenstates are not normalizable.

When we speak about the state of a particle and describe it with the wavefunction $\psi(x)$ we are using the “position” representation of the state. The momentum operator in the position representation is given by

$$\hat{p} \equiv \frac{\hbar}{i} \frac{\partial}{\partial x} \quad (6.35)$$

This is to say that acting on a wavefunction we have

$$\hat{p}\psi(x) = \frac{\hbar}{i} \frac{d\psi}{dx} . \quad (6.36)$$

Note that the commutation relation (6.27) is satisfied by the above definitions, as we can check acting on any wavefunction:

$$\begin{aligned} [\hat{x}, \hat{p}]\psi(x) &= (\hat{x}\hat{p} - \hat{p}\hat{x})\psi(x) \\ &= \hat{x}\hat{p}\psi(x) - \hat{p}\hat{x}\psi(x) \\ &= \hat{x} \frac{\hbar}{i} \frac{d\psi}{dx} - \hat{p}x\psi(x) \\ &= \frac{\hbar}{i} x \frac{d\psi}{dx} - \frac{\hbar}{i} \frac{d}{dx}(x\psi) \\ &= \frac{\hbar}{i} x \frac{d\psi}{dx} - \frac{\hbar}{i}\psi - \frac{\hbar}{i}x \frac{d\psi}{dx} \\ &= i\hbar\psi(x) . \end{aligned} \quad (6.37)$$

Since the wavefunction is arbitrary, we have verified that our explicit representation of the operators \hat{x} (by multiplication) and \hat{p} (by differentiation) satisfies the commutation relation $[\hat{x}, \hat{p}] = i\hbar$.

Eigenstates of the momentum operator exist and are not normalizable. Defining

$$\psi_p(x) \equiv \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}, \quad (6.38)$$

we readily confirm that

$$\hat{p}\psi_p(x) = \frac{\hbar}{i} \frac{\partial}{\partial x} \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} = p \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} = p\psi_p(x). \quad (6.39)$$

So $\psi_p(x)$ is a momentum eigenstate with momentum eigenvalue p . It is a plane wave.

The so-called momentum representation is mathematically described by Fourier transforms. The Fourier transform $\tilde{\psi}(p)$ of $\psi(x)$ is defined by

$$\tilde{\psi}(p) \equiv \int_{-\infty}^{\infty} dx \frac{e^{-ipx/\hbar}}{\sqrt{2\pi\hbar}} \psi(x). \quad (6.40)$$

The function $\tilde{\psi}(p)$ encodes the same amount of information as $\psi(x)$. We call $\tilde{\psi}(p)$ the momentum space representation of the state. Clearly for each value of p , $\tilde{\psi}$ is a linear superposition of values of $\psi(x)$ for all x . We can view the Fourier transformation as a linear transformation, the action of a matrix that depends on p on the vector that represents $\psi(x)$. The inverse Fourier transform is written as

$$\psi(x) = \int_{-\infty}^{\infty} dp \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} \tilde{\psi}(p). \quad (6.41)$$

We can view this formula as an expansion of $\psi(x)$ in a basis of momentum eigenstates, with $\tilde{\psi}(p)$ the expansion coefficients.

We have seen that $\psi(x)$ and $\tilde{\psi}(p)$ are just two different representations of the same state:

$$\psi(x) \longleftrightarrow \tilde{\psi}(p). \quad (6.42)$$

The arrow above is implemented by Fourier Transformation. Calculate now the action of $\frac{\hbar}{i} \frac{d}{dx}$ on (6.41)

$$\frac{\hbar}{i} \frac{d}{dx} \psi(x) = \frac{\hbar}{i} \frac{d}{dx} \int_{-\infty}^{\infty} dp \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} \tilde{\psi}(p) = \int_{-\infty}^{\infty} dp \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} p \tilde{\psi}(p). \quad (6.43)$$

In the language of (6.42) we write this as

$$\frac{\hbar}{i} \frac{d}{dx} \psi(x) \longleftrightarrow p \tilde{\psi}(p). \quad (6.44)$$

We see that the momentum operator, viewed as the action of $\frac{\hbar}{i} \frac{d}{dx}$ in coordinate space, is simply multiplication by p on momentum space wavefunctions $\tilde{\psi}$:

$$\hat{p} \tilde{\psi}(p) = p \tilde{\psi}(p). \quad (6.45)$$

This is, of course, perfectly analogous to the way that \hat{x} acts on position space wavefunctions.

Exercise. Verify that acting on momentum space wavefunctions the \hat{x} operator is represented by

$$\hat{x} \equiv i\hbar \frac{d}{dp}, \quad (\text{momentum representation}) \quad (6.46)$$

You can do this in two ways. Working with Fourier transforms, or by verifying (as in (6.37)) that it is consistent with $[\hat{x}, \hat{p}] = i\hbar$ acting on momentum space wavefunctions.

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

TWO STATE SYSTEMS

B. Zwiebach
November 15, 2013

Contents

1	Introduction	1
2	Spin precession in a magnetic field	2
3	The general two-state system viewed as a spin system	5
4	The ammonia molecule as a two-state system	7
5	Ammonia molecule in an electric field	11
6	Nuclear Magnetic Resonance	17

1 Introduction

A two-state system does not just have two states! It has two *basis* states, namely the state space is a two-dimensional complex vector space. For such a state space the Hamiltonian can be viewed as the most general Hermitian 2×2 matrix. In the case when the Hamiltonian is time-independent, this Hermitian matrix is characterized by four real numbers.

Two-state systems are idealizations that are valid when other degrees of freedom are ignored. A spin one-half particle is a two-state system with regards to spin, but being a particle, it may move and thus has position, or momentum degrees of freedom that imply a much larger, higher dimensional state space. Only if we ignore these degrees of freedom – perhaps because the particle is at rest – we can speak of a two-state system.

We will study here two specific examples of two-state systems. The first will be the ammonia molecule, which exhibits curious oscillations. The second will be spin one-half particles as used in nuclear magnetic resonance.

The mathematics of two-state systems is always the same. In fact any two-state system can be visualized as a spin system, and this will sometimes be quite useful.

2 Spin precession in a magnetic field

Let us first recall our earlier discussion of magnetic dipole moments. Classically we had the following relation valid for a charged particle

$$\boldsymbol{\mu} = \frac{q}{2m} \mathbf{S} \quad (2.1)$$

where $\boldsymbol{\mu}$ is the dipole moment, q is the charge of the particle, m its mass, and \mathbf{S} is its angular momentum, arising from its spinning. In the quantum world this equation gets modified by a constant unit-free factor g , different for each particle:

$$\boldsymbol{\mu} = g \frac{q}{2m} \mathbf{S} = g \frac{q\hbar}{2m} \frac{\mathbf{S}}{\hbar}. \quad (2.2)$$

Here $q\hbar/(2m)$ has the units of dipole moment. If we consider electrons and protons the following definitions are thus natural:

$$\begin{aligned} \text{Born magneton: } \mu_B &= \frac{e\hbar}{2m_e} = 5.78 \times 10^{-11} \frac{\text{MeV}}{\text{Tesla}}, \\ \text{Nuclear magneton: } \mu_N &= \frac{e\hbar}{2m_p} = 3.15 \times 10^{-14} \frac{\text{MeV}}{\text{Tesla}}. \end{aligned} \quad (2.3)$$

Note that the nuclear magneton is about two-thousand times smaller than the Bohr magneton. Nuclear magnetic dipole moments are much smaller than that of the electron! Including the g constant we have the following results. For an electron $g = 2$ and since the electron charge is negative we get

$$\boldsymbol{\mu}_e = -2 \mu_B \frac{\mathbf{S}}{\hbar}. \quad (2.4)$$

The dipole moment and the angular momentum are antiparallel. For a proton, the experimental result is

$$\boldsymbol{\mu}_p = 2.79 \mu_N \frac{\mathbf{S}}{\hbar}. \quad (2.5)$$

The neutron is neutral, so one would expect no magnetic dipole moment. But the neutron is not elementary: it is made by electrically charged quarks. A dipole moment is thus possible, depending on the way quarks are distributed. Indeed, experimentally,

$$\boldsymbol{\mu}_n = -1.91 \mu_N \frac{\mathbf{S}}{\hbar}. \quad (2.6)$$

Somehow the negative charge beats the positive charge in its contribution to the dipole moment of the neutron.

For notational convenience we introduce the constant γ from

$$\boldsymbol{\mu} = \gamma \mathbf{S}, \quad \text{with } \gamma = \frac{gq}{2m}.$$

(2.7)

If we insert the particle in a magnetic field \mathbf{B} , the Hamiltonian H_S for the spin system is

$$H_S = -\boldsymbol{\mu} \cdot \mathbf{B} = -\gamma \mathbf{B} \cdot \mathbf{S} = -\gamma (B_x \hat{S}_x + B_y \hat{S}_y + B_z \hat{S}_z). \quad (2.8)$$

If, for example we have a magnetic field $\vec{B} = B\hat{z}$ along the z axis, the Hamiltonian is

$$H_S = -\gamma B \hat{S}_z \quad (2.9)$$

The associated time evolution unitary operator is

$$\mathcal{U}(t, 0) = \exp\left(-\frac{iH_S t}{\hbar}\right) = \exp\left(-\frac{i(-\gamma B t)\hat{S}_z}{\hbar}\right) \quad (2.10)$$

We now recall a result that was motivated in the homework. You examined a unitary operator $R_{\mathbf{n}}(\alpha)$ defined by a unit vector \mathbf{n} and an angle α , and given by

$$R_{\mathbf{n}}(\alpha) = \exp\left(-\frac{i\alpha \hat{S}_{\mathbf{n}}}{\hbar}\right), \quad \text{with } \hat{S}_{\mathbf{n}} \equiv \mathbf{n} \cdot \mathbf{S}. \quad (2.11)$$

You found evidence that when acting on a spin state, this operator rotates it by an angle α about the axis defined by the vector \mathbf{n} . If we now compare (2.11) and (2.10) we conclude that $\mathcal{U}(t, 0)$ should generate a rotation by the angle $(-\gamma B t)$ about the z -axis. We now confirm this explicitly.

Consider a spin pointing at time equal zero along the direction specified by the angles (θ_0, ϕ_0) :

$$|\Psi, 0\rangle = \cos \frac{\theta_0}{2} |+\rangle + \sin \frac{\theta_0}{2} e^{i\phi_0} |-\rangle \quad (2.12)$$

Given the Hamiltonian $H_S = -\gamma B \hat{S}_z$ in (2.9) we have

$$H_S |\pm\rangle = \mp \frac{\gamma B \hbar}{2} \quad (2.13)$$

Then we have

$$\begin{aligned} |\Psi, t\rangle &= e^{-iH_S t/\hbar} |\Psi, 0\rangle = e^{-iH_S t/\hbar} \left(\cos \frac{\theta_0}{2} |+\rangle + \sin \frac{\theta_0}{2} e^{i\phi_0} |-\rangle \right) \\ &= \cos \frac{\theta_0}{2} e^{-iH_S t/\hbar} |+\rangle + \sin \frac{\theta_0}{2} e^{i\phi_0} e^{-iH_S t/\hbar} |-\rangle \\ &= \cos \frac{\theta_0}{2} e^{+i\gamma B t/2} |+\rangle + \sin \frac{\theta_0}{2} e^{i\phi_0} e^{-i\gamma B t/2} |-\rangle \end{aligned} \quad (2.14)$$

using (2.13). To recognize the resulting state it is convenient to factor out the phase that multiplies the $|+\rangle$ state:

$$|\Psi, t\rangle = e^{+i\gamma B t/2} \left(\cos \frac{\theta_0}{2} |+\rangle + \sin \frac{\theta_0}{2} e^{i(\phi_0 - \gamma B t)} |-\rangle \right). \quad (2.15)$$

Since the overall phase is not relevant we can now recognize the spin state as the state corresponding to the vector $\vec{n}(t)$ defined by angles

$$\begin{aligned}\theta(t) &= \theta_0, \\ \phi(t) &= \phi_0 - \gamma B t.\end{aligned}\tag{2.16}$$

Keeping θ constant while changing ϕ indeed corresponds to a rotation about the z axis and, after time t , the spin has rotated an angle $(-\gamma B t)$ as claimed above.

In fact spin states in a magnetic field precess in exactly the same way that magnetic dipoles in classical electromagnetism precess. The main fact from electromagnetic theory that we need is that in a magnetic field a dipole moment experiences a torque $\boldsymbol{\tau}$ given by

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}.\tag{2.17}$$

Then the familiar mechanics equation for the rate of change of angular momentum being equal to the torque gives

$$\frac{d\mathbf{S}}{dt} = \boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B} = \gamma \mathbf{S} \times \mathbf{B},\tag{2.18}$$

which we write as

$$\frac{d\mathbf{S}}{dt} = -\gamma \mathbf{B} \times \mathbf{S}.\tag{2.19}$$

We recognize that this equation states that the time dependent vector is rotating with angular velocity $\vec{\omega}_L$ given by

$$\boxed{\boldsymbol{\omega}_L = -\gamma \mathbf{B}.}\tag{2.20}$$

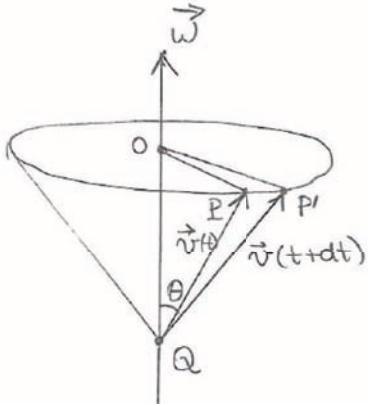
This is the so-called Larmor frequency. Indeed, this identification is standard in mechanics. A vector \mathbf{v} rotating with angular velocity $\boldsymbol{\omega}$ satisfies the differential equation

$$\frac{d\mathbf{v}}{dt} = \boldsymbol{\omega} \times \mathbf{v}.\tag{2.21}$$

You can convince yourself of this with the help of a simple picture (see Figure 1). Also note that the differential equation shows that the derivative of \mathbf{v} , given by the right-hand side, is orthogonal to \mathbf{v} because the cross product involves \mathbf{v} . This is as it should when the vector \mathbf{v} is rotated. Indeed, show that the above differential equation implies that $\frac{d}{dt}\mathbf{v} \cdot \mathbf{v} = 0$, so that the length of \mathbf{v} is unchanged.

The Hamiltonian of a general spin in a magnetic field (2.8) is then

$$\boxed{H_S = -\boldsymbol{\mu} \cdot \mathbf{B} = -\gamma \mathbf{B} \cdot \mathbf{S} = \boldsymbol{\omega}_L \cdot \mathbf{S}.}\tag{2.22}$$



$$\begin{aligned}
 \angle POP' &= \omega dt \\
 |OP| = |OP'| &= v \sin \theta \\
 |PP'| &= (\omega dt) v \sin \theta \\
 \vec{PP}' &= \vec{\omega} \times \vec{v} dt
 \end{aligned}$$

Figure 1: The vector $\mathbf{v}(t)$ and an instant later the vector $\mathbf{v}(t + dt)$. The angular velocity vector $\boldsymbol{\omega}$ is along the axis and \mathbf{v} rotates about is origin Q . At all times the vector \mathbf{v} and $\boldsymbol{\omega}$ make an angle θ . The calculations to the right should convince you that (2.21) is correct.

For time independent magnetic fields $\boldsymbol{\omega}_L$ is also time independent and the evolution operator is

$$\mathcal{U}(t, 0) = \exp(-iH_S t/\hbar) = \exp\left(-i \frac{\boldsymbol{\omega}_L \cdot \mathbf{S}}{\hbar} t\right). \quad (2.23)$$

If we write

$$\boldsymbol{\omega}_L = \omega_L \mathbf{n}, \quad \mathbf{n} \cdot \mathbf{n} = 1, \quad (2.24)$$

we have

$$\mathcal{U}(t, 0) = \exp\left(-i \frac{\omega_L t \hat{S}_{\mathbf{n}}}{\hbar}\right) = R_{\mathbf{n}}(\omega_L t), \quad (2.25)$$

where we compared with (2.11). The time evolution operator $\mathcal{U}(t, 0)$ rotates the spin states by the angle $\omega_L t$ about the \mathbf{n} axis. In other words

With $H_S = \boldsymbol{\omega}_L \cdot \mathbf{S}$ spin states precess with angular velocity $\boldsymbol{\omega}_L$.

3 The general two-state system viewed as a spin system

The most general time-independent Hamiltonian for a two-state system is a hermitian operator represented by the most general hermitian two-by-two matrix H . In doing so we are using some orthonormal basis $\{|1\rangle, |2\rangle\}$. In any such basis the matrix can be characterized by four real constants $g_0, g_1, g_2, g_3 \in \mathbb{R}$ as follows:

$$H = \begin{pmatrix} g_0 + g_3 & g_1 - ig_2 \\ g_1 + ig_2 & g_0 - g_3 \end{pmatrix} = g_0 \mathbf{1} + g_1 \sigma_1 + g_2 \sigma_2 + g_3 \sigma_3. \quad (3.27)$$

On the right-hand side we wrote the matrix as a sum of matrices, where $\mathbf{1}$ and the Pauli matrices σ_i , $i = 1, 2, 3$ are hermitian. We view (g_1, g_2, g_3) as a vector \mathbf{g} and then define

$$\mathbf{g} \cdot \boldsymbol{\sigma} \equiv g_1\sigma_2 + g_2\sigma_2 + g_3\sigma_3. \quad (3.28)$$

In this notation

$$H = g_0\mathbf{1} + \mathbf{g} \cdot \boldsymbol{\sigma}. \quad (3.29)$$

It is again convenient to introduce the magnitude g and the direction \mathbf{n} of \mathbf{g} :

$$\mathbf{g} = g\mathbf{n}, \quad \mathbf{n} \cdot \mathbf{n} = 1, \quad g = \sqrt{g_1^2 + g_2^2 + g_3^2}. \quad (3.30)$$

Now the Hamiltonian reads

$$H = g_0\mathbf{1} + g\mathbf{n} \cdot \boldsymbol{\sigma}. \quad (3.31)$$

Recall now that the spin states $|\mathbf{n}; \pm\rangle$ are eigenstates of $\mathbf{n} \cdot \boldsymbol{\sigma}$

$$\mathbf{n} \cdot \boldsymbol{\sigma} |\mathbf{n}; \pm\rangle = \pm |\mathbf{n}; \pm\rangle. \quad (3.32)$$

In writing the spin states, however, you must recall that what we call the z -up and z -down states are just the first and second basis states: $|+\rangle = |1\rangle$ and $|-\rangle = |2\rangle$. With this noted, the spin states $|\mathbf{n}; \pm\rangle$ are indeed the eigenstates of H , since using the last two equations above we have

$$H|\mathbf{n}; \pm\rangle = (g_0 \pm g)|\mathbf{n}; \pm\rangle. \quad (3.33)$$

This also shows that the energy eigenvalues are $g_0 \pm g$. In summary:

$$\text{Spectrum: } |\mathbf{n}; +\rangle \text{ with energy } g_0 + g, \quad |\mathbf{n}; -\rangle \text{ with energy } g_0 - g. \quad (3.34)$$

Thus our spin states allow us to write the general solution for the spectrum of the Hamiltonian (again, writing $|+\rangle = |1\rangle$ and $|-\rangle = |2\rangle$). Clearly the $|1\rangle$ and $|2\rangle$ states will generally have nothing to do with spin states. They are the basis states of any two-state system.

To understand the time evolution of states with the Hamiltonian H , we first rewrite H in terms of the spin operators, instead of the Pauli matrices, recalling that $\mathbf{S} = \frac{\hbar}{2}\boldsymbol{\sigma}$. Using (3.29) we find

$$H = g_0\mathbf{1} + \frac{2}{\hbar}\mathbf{g} \cdot \mathbf{S}. \quad (3.35)$$

Comparison with the spin Hamiltonian $H_S = \boldsymbol{\omega}_L \cdot \mathbf{S}$ shows that in the system described by H the states precess with angular velocity $\boldsymbol{\omega}$ given by

$$\boldsymbol{\omega} = \frac{2}{\hbar} \mathbf{g}. \quad (3.36)$$

Note that the part $g_0\mathbf{1}$ of the Hamiltonian H does not rotate states during time evolution; it simply multiplies states by the time-dependent phase $\exp(-ig_0t/\hbar)$.

Operationally, if H is known, the vector ω above is immediately calculable. And given a normalized state $\alpha|1\rangle + \beta|2\rangle$ of the system ($|\alpha|^2 + |\beta|^2 = 1$), we can identify the corresponding spin state $|\mathbf{n};+\rangle = \alpha|+\rangle + \beta|-\rangle$. The time evolution of the spin state is due to Larmor precession and is intuitively understood. With this result, the time evolution of the state in the original system is simply obtained by letting $|+\rangle \rightarrow |1\rangle$ and $|-\rangle \rightarrow |2\rangle$ in the precessing spin state.

4 The ammonia molecule as a two-state system

The ammonia molecule NH_3 is composed of four atoms, one nitrogen and three hydrogen. Ammonia is naturally a gas, without color, but with a pungent odor. It is mostly used for fertilizers, and also for cleaning products and pharmaceuticals.

The ammonia molecule takes the shape of a flattened tetrahedron. If we imagine the three hydrogen atoms forming an equilateral triangle at the base, the nitrogen atom sits atop. The angle formed between any two lines joining the nitrogen to the hydrogen is about 108° – this indeed corresponds to a flattened tetrahedron since a regular tetrahedron would have a 60° angle. If the nitrogen was pushed all the way down to the base, the angle would be 120° .

The ammonia molecule has electronic excitations, vibrational excitations and rotational excitations. Those must largely be ignored in the two-state description of the molecule. The two states arise from transitions in which the nitrogen atom flips from being above the fixed hydrogen plane to being below the hydrogen plane. Since such a flip could be mimicked by a full rotation of the molecule, we can describe the transition more physically by considering the molecule spinning about the axis perpendicular to the hydrogen plane, with the N up, where up is the direction of the angular momentum. The transition would have the N down, or against the angular momentum of the rotating molecule.

More briefly, the two states are: nitrogen up, or nitrogen down. Both are classically stable configurations separated by a potential energy barrier. In classical mechanics these are the two options and they are degenerate in energy.

As long as the energy barrier is not infinite, in quantum mechanics the degeneracy is broken. This, of course is familiar. We can roughly represent the potential experienced by the nitrogen atom as the potential $V(z)$ in figure 3, where the two equilibrium positions of the nitrogen are at $\pm z_0$ and they are separated by a large barrier. In such a potential the ground state, which is symmetric, and the first excited state, which is antisymmetric, are almost degenerate in energy when the barrier is high. If the potential barrier was infinite, the two possible eigenstates would be the nitrogen wavefunction localized about z_0 and the nitrogen wavefunction localized about $-z_0$. Moreover, those states would be degenerate in energy. But with a large but finite barrier the ground state is represented by a wavefunction $\psi_g(z)$ even in z , as shown below the potential.

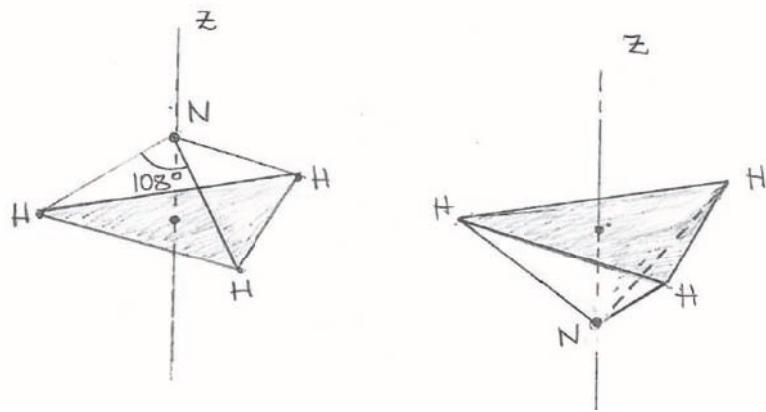


Figure 2: The ammonia molecule looks like a flattened tetrahedron. The nitrogen atom can be up or down with respect to the plane defined by the three hydrogen atoms. These are the two states of the ammonia molecule.

This even wavefunction is roughly the superposition, with the same sign, of the two localized wavefunctions. The next excited state $\psi_e(z)$ is odd in z and is roughly the superposition, this time with opposite signs, of the two localized wavefunctions.

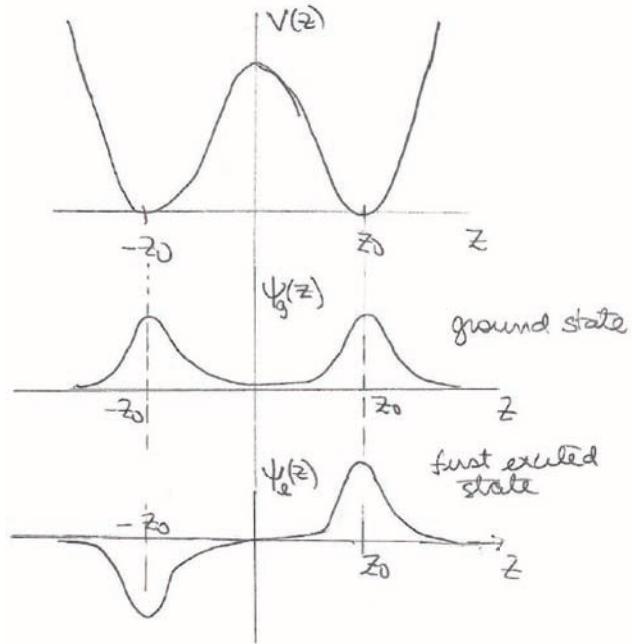


Figure 3: The potential $V(z)$ experienced by the nitrogen atom. There are two classically stable positions $\pm z_0$. The ground state and (first) excited state wavefunctions $\psi_g(z)$ and $\psi_e(z)$ are sketched below the potential.

Let us attempt a quantitative description of the situation. Let us label the two possible states:

$$|\uparrow\rangle \text{ is nitrogen up, } |\downarrow\rangle \text{ is nitrogen down} \quad (4.37)$$

We can associate to $|\uparrow\rangle$ a (positive) wavefunction localized around z_0 and to $|\downarrow\rangle$ a (positive) wavefunction localized around $-z_0$. Suppose the energy barrier is infinite. In this case the two states above must be energy eigenstates with the same energy E_0 :

$$\begin{aligned} H|\uparrow\rangle &= E_0|\uparrow\rangle, \\ H|\downarrow\rangle &= E_0|\downarrow\rangle. \end{aligned} \quad (4.38)$$

The energy E_0 is arbitrary and will not play an important role. Choosing a basis

$$|1\rangle \equiv |\uparrow\rangle, \quad |2\rangle \equiv |\downarrow\rangle, \quad (4.39)$$

the Hamiltonian, in this basis takes the form of the two-by-two matrix

$$H = \begin{pmatrix} E_0 & 0 \\ 0 & E_0 \end{pmatrix}. \quad (4.40)$$

The ability to tunnel must correspond to off-diagonal elements in the Hamiltonian matrix – there is no other option, in fact! So we must have a nonzero $H_{12} = \langle 1|H|2\rangle \neq 0$. Since the Hamiltonian must be hermitian, we must have $H_{12} = H_{21}^*$. For the time being we will take the off-diagonal elements to be real and therefore:

$$H_{12} = H_{21} = -\Delta, \quad \Delta > 0. \quad (4.41)$$

The sign of the real constant Δ is conventional. We could change it by a change of basis in which we let, for example $|2\rangle \rightarrow -|2\rangle$. Our choice will be convenient. The full Hamiltonian is now

$$H = \begin{pmatrix} E_0 & -\Delta \\ -\Delta & E_0 \end{pmatrix} = E_0 \mathbf{1} - \Delta \sigma_1, \quad (4.42)$$

where in the last step we wrote the matrix as a sum of a real number times the two-by-two identity matrix plus another real number times the first Pauli matrix. Both the identity matrix and the Pauli matrix are hermitian, consistent with having a hermitian Hamiltonian. The eigenvalues of the Hamiltonian follow from the equation

$$\begin{vmatrix} E_0 - \lambda & -\Delta \\ -\Delta & E_0 - \lambda \end{vmatrix} = 0 \quad \rightarrow \quad (E_0 - \lambda)^2 = \Delta^2 \quad \rightarrow \quad \lambda_{\pm} = E_0 \pm \Delta. \quad (4.43)$$

The eigenstates corresponding to these eigenvalues are

$$\begin{aligned} |G\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle), \quad E = E_0 - \Delta, \quad \text{Ground state} \\ |E\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle), \quad E = E_0 + \Delta, \quad \text{Excited state} \end{aligned} \quad (4.44)$$

Here $|G\rangle$ is the ground state (thus the G) and $|E\rangle$ is the excited state (thus the E). Our sign choices make the correspondence of the states with the wavefunction in figure 3 clear. The ground state $|G\rangle$ is a superposition, with the same sign, of the two localized (positive) wavefunctions. The excited state $|E\rangle$ has the wavefunction localized at z_0 (corresponding to $|1\rangle$) with a plus sign and the wavefunction localized at $-z_0$ (corresponding to $|2\rangle$) with a negative sign. Note that in the notation of section 3, the Hamiltonian in (4.42) corresponds to $\mathbf{g} = -\Delta \mathbf{x}$ and therefore $\mathbf{n} = -\mathbf{x}$ and $g = \Delta$. The excited state corresponds to the spin state $\frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)$, which points in the $-\mathbf{x}$ direction. The ground state corresponds to the spin state $\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$ which points in the $+\mathbf{x}$ direction.

The energy difference between these two eigenstates is 2Δ , which for the ammonia molecule takes the value

$$2\Delta = 0.9872 \times 10^{-4} \text{ ev}. \quad (4.45)$$

Setting this energy equal to the energy of a photon that may be emitted in such a transition we find

$$2\Delta = \hbar\omega = h\nu, \quad \nu = 23.870 \times 10^9 \text{ Hz} = 23.870 \text{ GHz}, \quad (4.46)$$

corresponding to a wavelength λ of

$$\lambda = 1.2559 \text{ cm}. \quad (4.47)$$

Let us consider the time evolution of an ammonia molecule that at $t = 0$ is in the state $|\uparrow\rangle$. Using (4.44) we express the initial state in terms of energy eigenstates:

$$|\Psi, 0\rangle = |\uparrow\rangle = \frac{1}{\sqrt{2}}(|G\rangle + |E\rangle). \quad (4.48)$$

The time-evolved state can now be readily written:

$$|\Psi, t\rangle = \frac{1}{\sqrt{2}}\left(e^{-i(E-\Delta)t/\hbar}|G\rangle + e^{-i(E+\Delta)t/\hbar}|E\rangle\right), \quad (4.49)$$

and we now rewrite it in terms of the $|\uparrow\rangle, |\downarrow\rangle$ states as follows

$$\begin{aligned} |\Psi, t\rangle &= \frac{1}{\sqrt{2}}\left(e^{-i(E-\Delta)t/\hbar}\frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) + e^{-i(E+\Delta)t/\hbar}\frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle)\right), \\ &= \frac{1}{2}e^{-iEt/\hbar}\left(e^{i\Delta t/\hbar}(|\uparrow\rangle + |\downarrow\rangle) + e^{-i\Delta t/\hbar}(|\uparrow\rangle - |\downarrow\rangle)\right), \\ &= e^{-iEt/\hbar}\left[\cos\left(\frac{\Delta t}{\hbar}\right)|\uparrow\rangle + i\sin\left(\frac{\Delta t}{\hbar}\right)|\downarrow\rangle\right]. \end{aligned} \quad (4.50)$$

The above time-dependent state oscillates from $|\uparrow\rangle$ to $|\downarrow\rangle$ with angular frequency $\omega = \Delta/\hbar \simeq 23 \text{ GHz}$. The probabilities $P_{\uparrow\downarrow}(t)$ that the state is found with nitrogen up or down are simply

$$\begin{aligned} P_{\uparrow}(t) &= |\langle \uparrow | \Psi, t \rangle|^2 = \cos^2\left(\frac{\Delta t}{\hbar}\right), \\ P_{\downarrow}(t) &= |\langle \downarrow | \Psi, t \rangle|^2 = \sin^2\left(\frac{\Delta t}{\hbar}\right). \end{aligned} \quad (4.51)$$

A plot of these is shown in Figure 4

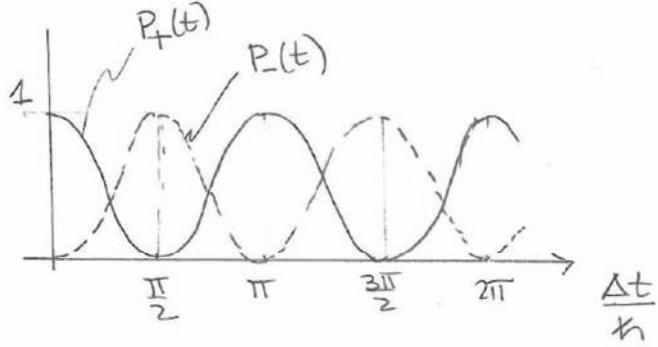


Figure 4: If the nitrogen atom starts in the state $|\uparrow\rangle$ at $t = 0$ (up position) then the probability $P_{\uparrow}(t)$ (shown as a continuous line) oscillates between one and zero. The probability $P_{\downarrow}(t)$ to be found in the state $|\downarrow\rangle$ is shown in dashed lines. Of course, $P_{\uparrow}(t) + P_{\downarrow}(t) = 1$. (Figure needs updating to change \pm to $\uparrow\downarrow$.)

5 Ammonia molecule in an electric field

Let us now consider the electrostatic properties of the ammonia molecule. The electrons tend to cluster towards the nitrogen, leaving the nitrogen vertex slightly negative and the hydrogen plane slightly positive. As a result we get an electric dipole moment μ that points down – when the nitrogen is up. The energy E of a dipole in an electric field \mathcal{E} is

$$E = -\mu \cdot \mathcal{E}. \quad (5.1)$$

With the electric field $\mathcal{E} = \mathcal{E}\mathbf{z}$ along the z axis, and $\mu = -\mu\mathbf{z}$, with $\mu > 0$, the state $|\uparrow\rangle$ with nitrogen up gets an extra positive contribution to the energy equal to $\mu\mathcal{E}$, while the $|\downarrow\rangle$ state gets the extra piece $-\mu\mathcal{E}$. The new Hamiltonian, including the effects of the electric field is then

$$H = \begin{pmatrix} E_0 + \mu\mathcal{E} & -\Delta \\ -\Delta & E_0 - \mu\mathcal{E} \end{pmatrix} = E_0 \mathbf{1} - \Delta \sigma_1 + \mu\mathcal{E} \sigma_3, \quad (5.2)$$

This corresponds to $g = \sqrt{(\mu\mathcal{E})^2 + \Delta^2}$ and therefore the energy eigenvalues are

$$\begin{aligned} E_E(\mathcal{E}) &= E_0 + \sqrt{\mu^2\mathcal{E}^2 + \Delta^2}, \\ E_G(\mathcal{E}) &= E_0 - \sqrt{\mu^2\mathcal{E}^2 + \Delta^2}, \end{aligned} \quad (5.3)$$

where we added the subscripts E for excited and G for ground to identify the energies as those of the excited and ground states when $\mathcal{E} = 0$. For small \mathcal{E} , or more precisely, small $\mu\mathcal{E}/\Delta$, we

have

$$\begin{aligned} E_E(\mathcal{E}) &\simeq E_0 + \Delta + \frac{\mu^2 \mathcal{E}^2}{2\Delta} + \mathcal{O}(\mathcal{E}^4), \\ E_G(\mathcal{E}) &\simeq E_0 - \Delta - \frac{\mu^2 \mathcal{E}^2}{2\Delta} + \mathcal{O}(\mathcal{E}^4), \end{aligned} \quad (5.4)$$

while for large $\mu\mathcal{E}$

$$\begin{aligned} E_E(\mathcal{E}) &\simeq E_0 + \mu\mathcal{E} + \mathcal{O}(1/\mathcal{E}), \\ E_G(\mathcal{E}) &\simeq E_0 - \mu\mathcal{E} + \mathcal{O}(1/\mathcal{E}). \end{aligned} \quad (5.5)$$

A plot of the energies is shown in Figure 5.

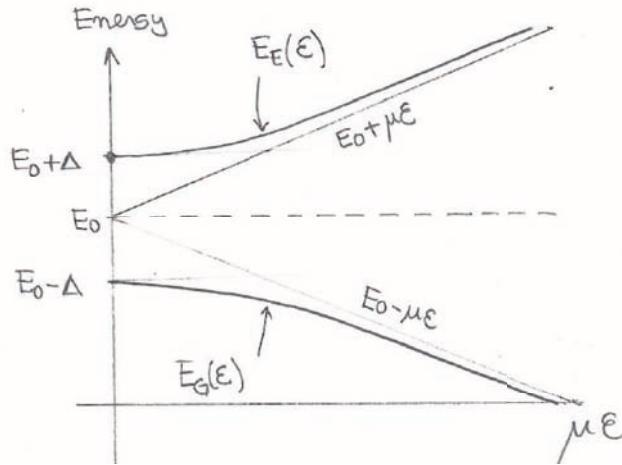


Figure 5: The energy levels of the two states of the ammonia molecule as a function of the magnitude \mathcal{E} of electric field. The higher energy state, with energy $E_E(\mathcal{E})$ coincides with $|E\rangle$ when $\mathcal{E} = 0$. The lower energy state, with energy $E_G(\mathcal{E})$ coincides with $|G\rangle$ when $\mathcal{E} = 0$.

This analysis gives us a way to split a beam of ammonia molecules into two beams, one with molecules in the state $|G\rangle$ and one with molecules in the state $|E\rangle$. As shown in the figure we have a beam entering a region with a spatially dependent electric field. The electric field gradient points up: the magnitude of the field is larger above than below. In a practical device $\mu\mathcal{E} \ll \Delta$ and we can use (5.4). A molecule in the $|E\rangle$ state will tend to go to the region of lower $|\mathcal{E}|$ as this is the region of low energy. Similarly a molecule in the $|G\rangle$ state will tend to go to the region of larger $|\mathcal{E}|$. Thus this device acts as a beam splitter.

The idea now is build a resonant electromagnetic cavity tuned to the frequency of 23.87 GHz and with very small losses (a high Q cavity). On one end, through a small hole, we let in a beam of ammonia molecules in the $|E\rangle$ state. These molecules exit the cavity through another

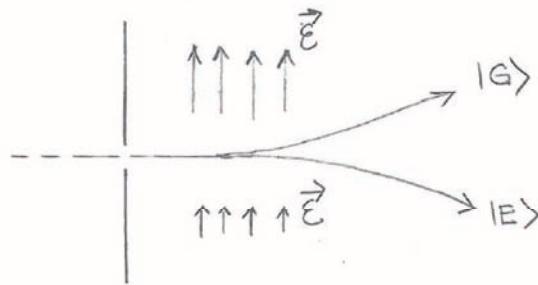


Figure 6: If a beam of ammonia molecules is exposed to an electric field with a strong gradient, molecules in the ground state $|G\rangle$ are deflected towards the stronger field (up) while molecules in the excited state $|E\rangle$ are deflected towards the weaker field (down).

hole on the opposite side (see Figure 7). If the design is done right, they exit on the ground state $|G\rangle$ thus having yielded an energy $2\Delta = \hbar\omega_0$ to the cavity. The ammonia molecules in the cavity interact with a spontaneously created electric field \mathcal{E} that oscillates with the resonant frequency. The interaction with such field induces the transition $|E\rangle \rightarrow |G\rangle$. This transition also feeds energy into the field. We want to understand this transition.

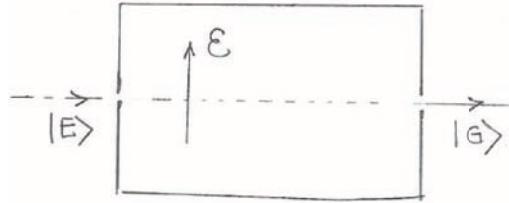


Figure 7: A resonant cavity tuned for 23.87GHz. A beam of ammonia molecules in the excited state $|E\rangle$ enter from the left. If properly designed, the molecules exit the cavity from the right on the ground state $|G\rangle$. In this process each molecule adds energy 2Δ to the electromagnetic field in the cavity.

The mathematics of the transition is clearer if we express the Hamiltonian in the primed basis

$$|1'\rangle \equiv |E\rangle, \quad |2'\rangle \equiv |G\rangle. \quad (5.6)$$

instead of the basis $|1\rangle = |\uparrow\rangle, |2\rangle = |\downarrow\rangle$ used to describe the Hamiltonian in (5.2). We can use

(4.44) to calculate the new matrix elements. For example

$$\begin{aligned}
\langle E|H|E\rangle &= \frac{1}{2}(\langle \uparrow | - \langle \downarrow |)H(|\uparrow\rangle - |\downarrow\rangle) \\
&= \frac{1}{2}(\langle \uparrow |H|\uparrow\rangle - \langle \uparrow |H|\downarrow\rangle - \langle \downarrow |H|\uparrow\rangle + \langle \downarrow |H|\downarrow\rangle) \\
&= \frac{1}{2}(E_0 + \mu\mathcal{E} - (-\Delta) - (-\Delta) + E_0 - \mu\mathcal{E}) \\
&= E_0 + \Delta,
\end{aligned} \tag{5.7}$$

and

$$\begin{aligned}
\langle E|H|G\rangle &= \frac{1}{2}(\langle \uparrow | - \langle \downarrow |)H(|+\rangle + |\downarrow\rangle) \\
&= \frac{1}{2}(\langle +|H|\uparrow\rangle + \langle \uparrow |H|\downarrow\rangle - \langle \downarrow |H|\uparrow\rangle - \langle \downarrow |H|\downarrow\rangle) \\
&= \frac{1}{2}(E_0 + \mu\mathcal{E} + (-\Delta) - (-\Delta) - (E_0 - \mu\mathcal{E})) \\
&= \mu\mathcal{E}.
\end{aligned} \tag{5.8}$$

and similarly $\langle G|H|G\rangle = E_0 - \Delta$. All in all the Hamiltonian in the new basis is given by

$$H = \begin{pmatrix} E_0 + \Delta & \mu\mathcal{E} \\ \mu\mathcal{E} & E_0 - \Delta \end{pmatrix} \quad \text{In the } |1'\rangle = |E\rangle, |2'\rangle = |G\rangle \text{ basis.} \tag{5.9}$$

We then write the wavefunction in terms of the amplitudes C_E and C_G to be in the $|E\rangle$ or $|G\rangle$ states respectively,

$$|\Psi\rangle = \begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix}. \tag{5.10}$$

The constant energy E_0 is not relevant – it can be set to any value and we choose the value zero. Doing so, the Schrödinger equation takes the form

$$i\hbar \frac{d}{dt} \begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix} = \begin{pmatrix} \Delta & \mu\mathcal{E} \\ \mu\mathcal{E} & -\Delta \end{pmatrix} \begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix}. \tag{5.11}$$

A strategy to solve this equation is to imagine that $\mu\mathcal{E}$ is very small compared to Δ so that

$$\begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix} = \begin{pmatrix} e^{-i\Delta t/\hbar} \beta_E(t) \\ e^{+i\Delta t/\hbar} \beta_G(t) \end{pmatrix}, \tag{5.12}$$

would be an exact solution with time-independent β_E and β_G if $\mu\mathcal{E} = 0$. When $\mu\mathcal{E}$ is small, we can expect solutions with β_E and β_G slowly varying in time (compared to the frequency Δ/\hbar) of the phases we have brought out to the open. We now substitute into (5.11), with the result (do the algebra!) with several terms canceling and

$$i\hbar \frac{d}{dt} \begin{pmatrix} \beta_E(t) \\ \beta_G(t) \end{pmatrix} = \begin{pmatrix} 0 & e^{i\omega_0 t} \mu\mathcal{E} \\ e^{-i\omega_0 t} \mu\mathcal{E} & 0 \end{pmatrix} \begin{pmatrix} \beta_E(t) \\ \beta_G(t) \end{pmatrix}, \quad \omega_0 \equiv \frac{2\Delta}{\hbar} \tag{5.13}$$

where we defined ω_0 as the frequency of a photon associated to the transition $|E\rangle \rightarrow |G\rangle$. This frequency is the resonant frequency of the cavity to be used. We now assume that in the cavity the electric field \mathcal{E} is at resonance so that

$$\mathcal{E}(t) = 2\mathcal{E}_0 \cos \omega_0 t = \mathcal{E}_0(e^{i\omega_0 t} + e^{-i\omega_0 t}) \quad (5.14)$$

so that

$$\begin{aligned} e^{i\omega_0 t} \mu \mathcal{E} &= \mu \mathcal{E}_0 (1 + e^{2i\omega_0 t}) \\ e^{-i\omega_0 t} \mu \mathcal{E} &= \mu \mathcal{E}_0 (1 + e^{-2i\omega_0 t}) \end{aligned} \quad (5.15)$$

We can now go back to the differential equation which gives

$$\begin{aligned} i \dot{\beta}_E(t) &= \frac{\mu \mathcal{E}_0}{\hbar} (1 + e^{2i\omega_0 t}) \beta_G(t) \\ i \dot{\beta}_G(t) &= \frac{\mu \mathcal{E}_0}{\hbar} (1 + e^{-2i\omega_0 t}) \beta_E(t). \end{aligned} \quad (5.16)$$

With $\mu \mathcal{E}_0$ small, the rates of change of β_E and β_G will necessarily be small, as $\mu \mathcal{E}_0$ appears multiplicatively on the right-hand side. Thus β_E and β_G are essentially constant during the period of oscillation of the exponential terms $e^{\pm 2i\omega_0 t}$. Since these exponentials have zero time-averaged values, they can be dropped. Thus, we get

$$\begin{aligned} \dot{\beta}_E(t) &= -i \frac{\mu \mathcal{E}_0}{\hbar} \beta_G(t) \\ \dot{\beta}_G(t) &= -i \frac{\mu \mathcal{E}_0}{\hbar} \beta_E(t). \end{aligned} \quad (5.17)$$

Taking another time derivative of the top equation we find;

$$\ddot{\beta}_E(t) = -\left(\frac{\mu \mathcal{E}_0}{\hbar}\right)^2 \beta_E(t) \quad (5.18)$$

This has the simple solution in terms of sines and cosines. If we assume that the molecule at time equal zero is indeed in the state $|E\rangle$ we then write

$$\beta_E(t) = \cos\left(\frac{\mu \mathcal{E}_0}{\hbar} t\right) \rightarrow \beta_G(t) = -i \sin\left(\frac{\mu \mathcal{E}_0}{\hbar} t\right). \quad (5.19)$$

The time dependent probability $P_E(t)$ to be in the $|E\rangle$ state is then

$$P_E(t) = |C_E(t)|^2 = \left|e^{-i\Delta t/\hbar} \beta_E(t)\right|^2 = \cos^2\left(\frac{\mu \mathcal{E}_0}{\hbar} t\right). \quad (5.20)$$

This is our desired result. The molecule that enters the cavity in the state $|E\rangle$ will leave the cavity in the state $|G\rangle$ if the travel time T is such that the probability $P_E(T)$ to be in $|E\rangle$ vanishes. For this we need

$$\cos\left(\frac{\mu \mathcal{E}_0}{\hbar} T\right) = 0 \rightarrow \frac{\mu \mathcal{E}_0}{\hbar} T = \frac{\pi}{2}, \frac{3\pi}{2}, \dots \quad (5.21)$$

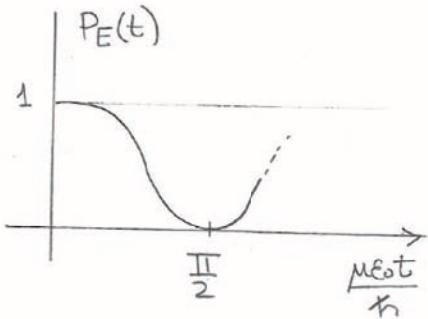


Figure 8: An ammonia molecule enters the resonant cavity at $t = 0$ in the excited state $|E\rangle$. We show probability $P_E(t)$ for the ammonia molecule to be in the excited $|E\rangle$ at time t . If the molecule exits at a time T for which $P_E(T) = 0$, the molecule will be in the ground state $|G\rangle$, as desired.

See figure 8.

If the velocity of the ammonia molecules is adjusted for this to happen, each molecule gives energy 2Δ to the cavity's electromagnetic field. The cavity's EM field, by producing the transition $|E\rangle \rightarrow |G\rangle$ of the traveling molecules, stimulates the emission of radiation. Moreover the energy released is in a field with the same configuration and frequency as the stimulating EM field. The molecules thus help build a *coherent* field in the cavity. Such a cavity is then a MASER, an acronym that stands for Microwave Amplification by Stimulated Emission of Radiation. The molecules are uncharged and therefore their motions produce no unwanted EM signals – this is in contrast to electrons in vacuum amplifiers, which produce *shot* noise.

Charles H. Townes, James P. Gordon, and H. J. Zeiger built the first ammonia maser working at Columbia University in 1953. As stated by Charles H. Townes in his Nobel lecture on December 11, 1964, “masers yield the most perfect amplification allowed by the uncertainty principle”. For such EM waves the uncertainty principle can be written as

$$\Delta n \Delta \phi \geq \frac{1}{2},$$

where Δn is the uncertainty in the number of photons in the field and $\Delta \phi$ is the phase uncertainty of the wave, in radians. For a coherent field $\Delta n = \sqrt{\bar{n}}$, with \bar{n} the expected number of photons. The saturation of the uncertainty principle leads to a phase uncertainty

$$\Delta \phi = \frac{1}{2\sqrt{\bar{n}}}, \quad (5.22)$$

that for any realistic \bar{n} is fantastically small.

6 Nuclear Magnetic Resonance

The problem we want to discuss is that of a spin in a time dependent magnetic field. This magnetic field has a time-independent z component and a circularly polarized field representing a magnetic field rotating on the (x, y) plane. More concretely, we have

$$\mathbf{B}(t) = B_0 \mathbf{z} + B_1 (\mathbf{x} \cos \omega t - \mathbf{y} \sin \omega t). \quad (6.23)$$

Typically, the constant z -component B_0 is larger than B_1 , the magnitude of the RF (radio-frequency) signal. The time dependent part of the field points along the x axis at $t = 0$ and is rotating with angular velocity ω in the clockwise direction of the (x, y) plane. The spin Hamiltonian is

$$H_S(t) = -\gamma \mathbf{B}(t) \cdot \mathbf{S} = -\gamma \left(B_0 \hat{S}_z + B_1 (\cos \omega t \hat{S}_x - \sin \omega t \hat{S}_y) \right). \quad (6.24)$$

Not only is this Hamiltonian time dependent, but the Hamiltonian at different times do not commute. So this is a nontrivial time evolution problem!

We attempt to simplify the problem by considering a frame of reference that rotates with the magnetic field. For this, imagine first the case when $H_S = 0$ because the magnetic field is zero. With no magnetic field spin states would simply be static. What would the Hamiltonian be in the frame rotating about the z -axis with angular frequency ω , just like the magnetic field above? In that frame, the spin states that are fixed in the original frame would be seen to rotate with positive angular velocity ω about the z direction. There must be a Hamiltonian that does have that effect. Since the unitary operator \mathcal{U} that generates this rotation is

$$\mathcal{U}(t) = \exp\left(-\frac{i\omega t \hat{S}_z}{\hbar}\right) \rightarrow H_{\mathcal{U}} = \omega \hat{S}_z, \quad (6.25)$$

where the expression $H_{\mathcal{U}}$ for the Hamiltonian in the rotating frame is read from the relation $\mathcal{U} = \exp(-iH_{\mathcal{U}}t/\hbar)$. For the original case, when the original Hamiltonian in the static frame is H_S , we will use the above operator \mathcal{U} to define a new rotating-frame state $|\Psi_R\rangle$ as follows

$$|\Psi_R, t\rangle \equiv \mathcal{U}(t)|\Psi, t\rangle. \quad (6.26)$$

If we knew $|\Psi_R, t\rangle$ we would know $|\Psi, t\rangle$. We wish to find out if the Schrödinger equation for $|\Psi_R\rangle$ becomes simpler. For this we must determine the corresponding Hamiltonian H_R . One quick way to do this is to note that the above equation implies that

$$|\Psi_R, t\rangle \equiv \mathcal{U}(t)\mathcal{U}_S(t)|\Psi, 0\rangle. \quad (6.27)$$

where we have introduced the unitary operator $\mathcal{U}_S(t)$ associated with the Hamiltonian $H_S(t)$ that evolves $|\Psi\rangle$ in time. Since the Hamiltonian associated to an arbitrary unitary time-evolution operator \mathcal{U} is $i\hbar(\partial_t \mathcal{U})\mathcal{U}^\dagger$ (if you don't recall this, derive it from a Schrödinger equation)

we have

$$\begin{aligned}
H_R &= i\hbar \partial_t (\mathcal{U} \mathcal{U}_S) \mathcal{U}_S^\dagger \mathcal{U}^\dagger \\
&= i\hbar (\partial_t \mathcal{U}) \mathcal{U}^\dagger + \mathcal{U} i\hbar (\partial_t \mathcal{U}_S) \mathcal{U}_S^\dagger \mathcal{U}^\dagger \\
\rightarrow H_R &= H_{\mathcal{U}} + \mathcal{U} H_S \mathcal{U}^\dagger,
\end{aligned} \tag{6.28}$$

where $H_{\mathcal{U}}$ is the Hamiltonian associated to \mathcal{U} . Since \mathcal{U} is the one above, we have $H_{\mathcal{U}} = \omega \hat{S}_z$ and therefore,

$$\begin{aligned}
H_R &= \omega \hat{S}_z + \exp\left(-\frac{i\omega t \hat{S}_z}{\hbar}\right) \left[-\gamma \left(B_0 \hat{S}_z + B_1 (\cos \omega t \hat{S}_x - \sin \omega t \hat{S}_y) \right) \right] \exp\left(\frac{i\omega t \hat{S}_z}{\hbar}\right) \\
&= (-\gamma B_0 + \omega) \hat{S}_z - \gamma B_1 \underbrace{\exp\left(-\frac{i\omega t \hat{S}_z}{\hbar}\right) (\cos \omega t \hat{S}_x - \sin \omega t \hat{S}_y) \exp\left(\frac{i\omega t \hat{S}_z}{\hbar}\right)}_{\hat{M}(t)}
\end{aligned} \tag{6.29}$$

and the big question is what is $\hat{M}(t)$. We can proceed by calculating the time-derivative of \hat{M} :

$$\begin{aligned}
\partial_t \hat{M} &= e^{-\frac{i\omega t \hat{S}_z}{\hbar}} \left(-\frac{i\omega}{\hbar} [\hat{S}_z, \cos \omega t \hat{S}_x - \sin \omega t \hat{S}_y] + [-\omega \sin \omega t \hat{S}_x - \omega \cos \omega t \hat{S}_y] \right) e^{\frac{i\omega t \hat{S}_z}{\hbar}} \\
&= e^{-\frac{i\omega t \hat{S}_z}{\hbar}} \left(-\frac{i\omega}{\hbar} [i\hbar \cos \omega t \hat{S}_y + i\hbar \sin \omega t \hat{S}_x] + [-\omega \sin \omega t \hat{S}_x - \omega \cos \omega t \hat{S}_y] \right) e^{\frac{i\omega t \hat{S}_z}{\hbar}} \\
&= e^{-\frac{i\omega t \hat{S}_z}{\hbar}} \left(\omega \cos \omega t \hat{S}_y + \omega \sin \omega t \hat{S}_x - \omega \sin \omega t \hat{S}_x - \omega \cos \omega t \hat{S}_y \right) e^{\frac{i\omega t \hat{S}_z}{\hbar}} \\
&= 0.
\end{aligned} \tag{6.30}$$

This is very good news. Since \hat{M} has no time dependence, we can evaluate it at any time. The simplest time is $t = 0$ and we thus find that

$$\hat{M}(t) = \hat{S}_x. \tag{6.31}$$

As a result we have a totally time-independent H_R :

$$\begin{aligned}
H_R &= (-\gamma B_0 + \omega) \hat{S}_z - \gamma B_1 \hat{S}_x \\
&= -\gamma \left[\left(B_0 - \frac{\omega}{\gamma} \right) \hat{S}_z + B_1 \hat{S}_x \right] \\
&= -\gamma \left[B_0 \left(1 - \frac{\omega}{\omega_0} \right) \hat{S}_z + B_1 \hat{S}_x \right],
\end{aligned} \tag{6.32}$$

using $\omega_0 = \gamma B_0$ for the Larmor frequency associated with the constant component of the field. We thus have a Hamiltonian H_R that can be associated with a magnetic field \mathbf{B}_R given by Setting

$$H_R = -\gamma \mathbf{B}_R \cdot \mathbf{S}, \quad \rightarrow \quad \mathbf{B}_R = B_1 \mathbf{x} + B_0 \left(1 - \frac{\omega}{\omega_0} \right) \mathbf{z}. \tag{6.33}$$

The full solution for the state is obtained beginning with (6.26) and (6.25):

$$|\Psi, t\rangle = \mathcal{U}^\dagger(t)|\Psi_R, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right]|\Psi_R, t\rangle. \quad (6.34)$$

Since H_R is a time-independent Hamiltonian, we have that the full time evolution is

$$|\Psi, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right] \exp\left[-i\frac{(-\gamma \mathbf{B}_R \cdot \mathbf{S})t}{\hbar}\right] |\Psi, 0\rangle \quad (6.35)$$

where the solution for $|\Psi_R, t\rangle$ is simply the time evolution from the $t = 0$ state generated by the Hamiltonian H_R . We have thus found that

$$|\Psi, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right] \exp\left[i\frac{\gamma \mathbf{B}_R \cdot \mathbf{S} t}{\hbar}\right] |\Psi, 0\rangle. \quad (6.36)$$

Exercise. Verify that for $B_1 = 0$ the above solution reduces to the one describing precession about the z -axis.

In the applications to be discussed below we always have

$$B_1 \ll B_0. \quad (6.37)$$

Consider now the evolution of a spin that initially points in the positive z direction. We look at two cases:

- $\omega \ll \omega_0$. In this case from (6.33) we have

$$\mathbf{B}_R \simeq B_0 \mathbf{z} + B_1 \mathbf{x}. \quad (6.38)$$

This is a field mostly along the z axis, but tipped a little towards the x axis. The right-most exponential in (6.36) makes the spin precess about the direction \mathbf{B}_R quite quickly, for $|\mathbf{B}_R| \sim B_0$ and thus the angular rate of precession is pretty much ω_0 . The next exponential in (6.36) induces a rotation about the z -axis with smaller angular velocity ω . This is shown in the figure.

- $\omega = \omega_0$. In this case from (6.33) we have

$$\mathbf{B}_R = B_1 \mathbf{x}. \quad (6.39)$$

In this case the right-most exponential in (6.36) makes the spin precess about the x axis it goes from the z axis towards the y axis with angular velocity $\omega_1 = \gamma B_1$. If we time the RF signal to last a time T such that

$$\omega_1 T = \frac{\pi}{2} \quad \rightarrow \quad T = \frac{\pi}{2\gamma B_1}, \quad (6.40)$$

the state $|\Psi_R, T\rangle$ points along the y axis. The effect of the other exponential in (6.36) is just to rotate the spin on the (x, y) plane. We have

$$|\Psi, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right] |\Psi_R, t\rangle, \quad t < T, \quad (6.41)$$

and if the RF pulse turns off after time T ,

$$|\Psi, t\rangle = \exp\left[\frac{i\omega t \hat{S}_z}{\hbar}\right] |\Psi_R, T\rangle, \quad t > T, \quad (6.42)$$

The state $|\Psi, t\rangle$ can be visualized as a spin that is slowly rotating with angular velocity ω_1 from the z axis towards the plane, while rapidly rotating around the z axis with angular velocity ω . As a result the tip of the spin is performing a spiral motion on the surface of a hemisphere. By the time the polar angle reaches $\pi/2$ the RF signal turns off and the spin now just rotates within the (x, y) plane. This is called a 90° pulse.

Magnetic Resonance Imaging Based on work on nuclear magnetic resonance by Edward Purcell, who worked at MIT's radiation laboratory, and Felix Bloch. They got the Nobel prize for this work in 1952.

The NMR work led to the development of the technique called MRI for Magnetic Resonance Imaging. First studies in humans was done in 1977. The new perspective, as compared to X-rays, was that MRI allows one to distinguish various soft tissues.

The human body is mostly composed of water molecules. In those we have many hydrogen atoms, whose nuclei are protons and are the main players through their magnetic dipole moments (nuclear dipoles).

With a given external and large constant magnetic field B_0 , at a given temperature, there is a net alignment of nuclear spins along B_0 . This is called the “longitudinal magnetization”. For all intents and purposes we have a net number of spins in play.

We apply a 90° pulse so that we get the spins to rotate with Larmor frequency ω_0 in the (x, y) plane. These rotating dipoles produce an oscillating magnetic field which is a signal that can be picked up by a receiver. The magnitude of the signal is proportional to the proton density. This is the first piece of information and allows differentiation of tissues.

The above signal from the rotation of the spins decays with a time constant T_2 that can be measured and is typically much smaller than a second. This decay is attributed to interactions between the spins. A T_2 weighted image allows doctors to detect abnormal accumulation of fluids (edema).

There is another time constant T_1 (of order one second) that controls the time to regain the longitudinal magnetization. This effect is due to the spins interacting with the rest of

the lattice of atoms. White matter, grey matter, and cerebrospinal fluids are distinguished by different T_1 constants (they have about the same proton density).

MRI's commonly include the use of contrast agents, which are substances that shorten the time constant T_1 and are usually administered by injection into the blood stream. The contrast agent (gadolinium) can accumulate at organs or locations where information is valuable. For a number of substances one can use the MRI apparatus to determine their (T_1, T_2) constants and build a table of data. This table can then be used as an aid to identify the results of MRI's.

The typical MRI machine has a B_0 of about 2T (two tesla or 20,000 gauss). This requires a superconducting magnet with cooling by liquid helium. For people with claustrophobia there are “open” MRI scanners that work with lower magnetic fields. In addition there are about three *gradient* magnets, each of about 200 gauss. They change locally the value of B_0 and thus provide spatial resolution as the Larmor frequency becomes spatially dependent. One can thus attain spatial resolutions of about half a millimiter! MRI's are considered safe, as there is no evidence of biological harm caused by very large static magnetic fields.

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Chapter 1: Key Features of Quantum Mechanics

Quantum mechanics is now almost one-hundred years old, but we are still discovering some of its surprising features and it remains the subject of much investigation and speculation. The framework of quantum mechanics is a rich and elegant extension of the framework of classical physics. It is also counterintuitive and almost paradoxical.

Quantum physics has replaced classical physics as the correct fundamental description of our physical universe. It is used routinely to describe most phenomena that occur at short distances. Quantum physics is the result of applying the framework of quantum mechanics to different physical phenomena. We thus have Quantum Electrodynamics, when quantum mechanics is applied to electromagnetism, Quantum Optics, when it is applied to light and optical devices, or Quantum Gravity, when it is applied to gravitation. Quantum mechanics indeed provides a remarkably coherent and elegant framework. The era of quantum physics begins in 1925, with the discoveries of Schrödinger and Heisenberg. The seeds for these discoveries were planted by Planck, Einstein, Bohr, de Broglie, and others. It is a tribute to human imagination that we have been able to discover the counterintuitive and abstract set of rules that define quantum mechanics. Here we aim to explain and provide some perspective on the main features of this framework.

We will begin by discussing the property of linearity, which quantum mechanics shares with electromagnetic theory. This property tells us what kind of theory quantum mechanics is and why, it could be argued, it is simpler than classical mechanics. We then turn to photons, the particles of light. We use photons and polarizers to explain why quantum physics is not deterministic and, in contrast with classical physics, the results of some experiments cannot be predicted. Quantum mechanics is a framework in which we can only predict the *probabilities* for the various outcomes of any given experiment. Our next subject is quantum superpositions, in which a quantum object somehow manages to exist simultaneously in two mutually incompatible states. A quantum light-bulb, for example, could be in a state in which it is both on and off at the same time!

1 Linearity of the equations of motion

In physics a theory is usually described by a set of equations for some quantities called the **dynamical variables** of the theory. After writing a theory, the most important task is finding solutions of the equations. A solution of the equations describes a possible reality, according to the theory. Because an expanding universe is a solution of Albert Einstein's gravitational equations, for example, it follows that an expanding universe is possible, according to this theory. A single theory may have many solutions, each describing a possible reality.

There are linear theories and nonlinear theories. Nonlinear theories are more complex than linear theories. In a linear theory a remarkable fact takes place: if you have two solutions you obtain a third solution of the theory simply by adding the two solutions. An example of a beautiful linear theory is Maxwell's theory of electromagnetism, a theory that governs the behavior of electric and magnetic fields. A field, as you probably know, is a quantity whose values may depend on position and on time. A simple solution of this theory describes an electromagnetic wave propagating in a given direction. Another simple solution could describe an electromagnetic wave propagating in a

different direction. Because the theory is linear, having the two waves propagating simultaneously, each in its own direction and without affecting each other, is a new and consistent solution. The sum is a solution in the sense that the electric field in the new solution is the sum of the electric field in the first solution plus the electric field in the second solution. The same goes for the magnetic field: the magnetic field in the new solution is the sum of the magnetic field in the first solution plus the magnetic field in the second solution. In fact you can add any number of solutions to still find a solution. Even if this sounds esoteric, you are totally familiar with it. The air around you is full of electromagnetic waves, each one propagating oblivious to the other ones. There are the waves of thousands of cell phones, the waves carrying hundreds of wireless internet messages, the waves from a plethora of radio-stations, TV stations, and many, many more. Today, a single transatlantic cable can carry simultaneously millions of telephone calls, together with huge amounts video and internet data. All of that courtesy of linearity.

More concretely, we say that Maxwell's equations are **linear** equations. A solution of Maxwell's equation is described by an electric field \mathbf{E} , a magnetic field \mathbf{B} , a charge density ρ and a current density \mathbf{J} , all collectively denoted as $(\mathbf{E}, \mathbf{B}, \rho, \mathbf{J})$. This collection of fields and sources satisfy Maxwell's equations. Linearity implies that if $(\mathbf{E}, \mathbf{B}, \rho, \mathbf{J})$ is a solution so is $(\alpha\mathbf{E}, \alpha\mathbf{B}, \alpha\rho, \alpha\mathbf{J})$, where all fields and sources have been multiplied by the constant α . Given two solutions

$$(\mathbf{E}_1, \mathbf{B}_1, \rho_1, \mathbf{J}_1), \quad \text{and} \quad (\mathbf{E}_2, \mathbf{B}_2, \rho_2, \mathbf{J}_2), \quad (1.1)$$

linearity also implies that we can obtain a new solution by adding them

$$(\mathbf{E}_1 + \mathbf{E}_2, \mathbf{B}_1 + \mathbf{B}_2, \rho_1 + \rho_2, \mathbf{J}_1 + \mathbf{J}_2). \quad (1.2)$$

The new solution may be called the superposition of the two original solutions.

It is not hard to explain what is, in general, a linear equation or a linear set of equations. Consider the equation

$$L u = 0, \quad (1.3)$$

where, schematically, u denotes the unknown. The unknown may be a number, or a function of time, a function of space, a function of time and space, essentially anything unknown! In fact, u could represent a collection of unknowns, in which case we would replace u above by u_1, u_2, \dots . The symbol L denotes a **linear operator**, an object that satisfies the following two properties

$$L(u_1 + u_2) = Lu_1 + Lu_2, \quad L(a u) = aLu, \quad (1.4)$$

where a is a number. Note that these conditions imply that

$$L(\alpha u_1 + \beta u_2) = \alpha Lu_1 + \beta Lu_2, \quad (1.5)$$

showing that if u_1 is a solution ($Lu_1 = 0$) and u_2 is a solution ($Lu_2 = 0$) then $\alpha u_1 + \beta u_2$ is also a solution. We call $\alpha u_1 + \beta u_2$ the **general superposition** of the solutions u_1 and u_2 . An example may help. Consider the equation

$$\frac{du}{dt} + \frac{1}{\tau} u = 0, \quad (1.6)$$

where τ is a constant with units of time. This is, in fact, a linear differential equation, and takes the form $L u = 0$ if we define

$$L u \equiv \frac{du}{dt} + \frac{1}{\tau} u \quad (1.7)$$

Exercise 1. Verify that (1.7) satisfies the conditions for a linear operator.

Einstein's theory of general relativity is a nonlinear theory whose dynamical variable is a gravitational field, the field that describes, for example, how planets move around a star. Being a nonlinear theory, you simply cannot add the gravitational fields of different solutions to find a new solution. This makes Einstein's theory rather complicated, by all accounts much more complicated than Maxwell theory. In fact, classical mechanics, as invented mostly by Isaac Newton, is also a nonlinear theory! In classical mechanics the dynamical variables are positions and velocities of particles, acted by forces. There is no general way to use two solutions to build a third.

Indeed, consider the equation of motion for a particle on a line under the influence of a time-independent potential $V(x)$, which is in general an arbitrary function of x . The dynamical variable in this problem is $x(t)$, the position as a function of time. Letting V' denote the derivative of V with respect to its argument, Newton's second law takes the form

$$m \frac{d^2x(t)}{dt^2} = -V'(x(t)). \quad (1.8)$$

The left-hand side is the mass times acceleration and the right hand side is the force experienced by the particle in the potential. It is probably worth to emphasize that the right hand side is the function $V'(x)$ evaluated for x set equal to $x(t)$:

$$V'(x(t)) \equiv \left. \frac{\partial V(x)}{\partial x} \right|_{x=x(t)}. \quad (1.9)$$

While we could have used here an ordinary derivative, we wrote a partial derivative as is commonly done for the general case of time dependent potentials. The reason equation (1.8) is not a linear equation is that the function $V'(x)$ is not linear. In general, for arbitrary functions u and v we expect

$$V'(au) \neq aV'(u), \quad \text{and} \quad V'(u+v) \neq V'(u) + V(v). \quad (1.10)$$

As a result given a solution $x(t)$, the scaled solution $\alpha x(t)$ is not expected to be a solution. Given two solutions $x_1(t)$ and $x_2(t)$ then $x_1(t) + x_2(t)$ is not guaranteed to be a solution either.

Exercise. What is the most general potential $V(x)$ for which the equation of motion for $x(t)$ is linear?

Quantum mechanics is a linear theory. The signature equation in this theory, the so-called Schrödinger equation is a linear equation for a quantity called the **wavefunction** and it determines its time evolution. The wavefunction is the dynamical variable in quantum mechanics but, curiously, its physical interpretation was not clear to Erwin Schrödinger when he wrote the equation in 1925. It was Max Born, who months later suggested that the wavefunction encodes probabilities. This was the correct physical interpretation, but it was thoroughly disliked by many, including Schrödinger, who remained unhappy about it for the rest of his life. The linearity of quantum mechanics implies a profound simplicity. In some sense quantum mechanics is simpler than classical mechanics. In quantum mechanics solutions can be added to form new solutions.

The wavefunction Ψ depends on time and may also depend on space. The Schrödinger equation (SE) is a partial differential equation that takes the form

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi, \quad (1.11)$$

where the Hamiltonian (or energy operator) \hat{H} is a linear operator that can act on wavefunctions:

$$\hat{H}(a\Psi) = a\hat{H}\Psi, \quad \hat{H}(\Psi_1 + \Psi_2) = \hat{H}(\Psi_1) + \hat{H}(\Psi_2), \quad (1.12)$$

with a a constant that in fact need not be real; it can be a complex number. Of course, \hat{H} itself does not depend on the wavefunction! To check that the Schrödinger equation is linear we cast it in the form $L\Psi = 0$ with L defined as

$$L\Psi \equiv i\hbar \frac{\partial \Psi}{\partial t} - \hat{H}\Psi \quad (1.13)$$

It is now a simple matter to verify that L is a linear operator. Physically this means that if Ψ_1 and Ψ_2 are solutions to the Schrödinger equation, then so is the superposition $\alpha\Psi_1 + \beta\Psi_2$, where α and β are both complex numbers, i.e. ($\alpha, \beta \in \mathbb{C}$)

2 Complex Numbers are Essential

Quantum mechanics is the first physics theory that truly makes use of *complex* numbers. The numbers most of us use for daily life (integers, fractions, decimals) are *real* numbers. The set of complex numbers is denoted by \mathbb{C} and the set of real numbers is denoted by \mathbb{R} . Complex numbers appear when we combine real numbers with the imaginary unit i , defined to be equal to the square root of minus one: $i \equiv \sqrt{-1}$. Being the square root of minus one, it means that i squared must give minus one: $i^2 = -1$. Complex numbers are fundamental in mathematics. An equation like $x^2 = -4$, for an unknown x cannot be solved if x has to be real. No real number squared gives you minus one. But if we allow for complex numbers, we have the solutions $x = \pm 2i$. Mathematicians have shown that all polynomial equations can be solved in terms of complex numbers.

A complex number z , in all generality, is a number of the form

$$z = a + ib \in \mathbb{C}, \quad a, b \in \mathbb{R}. \quad (2.1)$$

Here a and b are real numbers, and ib denotes the product of i with b . The number a is called the real part of z and b is called the imaginary part of z :

$$\operatorname{Re} z = a, \quad \operatorname{Im} z = b. \quad (2.2)$$

The complex conjugate z^* of z is defined by

$$z^* = a - ib. \quad (2.3)$$

You can quickly verify that a complex number z is real if $z^* = z$ and it is purely imaginary if $z^* = -z$. For any complex number $z = a + ib$ one can define the *norm* $|z|$ of the complex number to be a *positive, real* number given by

$$|z| = \sqrt{a^2 + b^2}. \quad (2.4)$$

You can quickly check that

$$|z|^2 = zz^*, \quad (2.5)$$

where $z^* \equiv a - ib$ is called the complex conjugate of $z = a + ib$. Complex numbers are represented as vectors in a two dimensional “complex plane”. The real part of the complex number is the x component of the vector and the imaginary part of the complex number is the y component. If you consider the unit length vector in the complex plane making an angle θ with the x axis has x component $\cos \theta$ and

y component $\sin \theta$. The vector is therefore the complex number $\cos \theta + i \sin \theta$. Euler's identity relates this to the exponential of $i\theta$:

$$e^{i\theta} = \cos \theta + i \sin \theta. \quad (2.6)$$

A complex number of the form $e^{i\chi}$, with χ real is called a *pure phase*.

While complex numbers are sometimes useful in classical mechanics or Maxwell theory, they are not strictly needed. None of the dynamical variables, which correspond to measurable quantities, is a complex number. In fact, complex numbers can't be measured at all: all measurements in physics result in real numbers. In quantum mechanics, however, complex numbers are fundamental. The Schrödinger equation involves complex numbers. Even more, the wavefunction, the dynamical variable of quantum mechanics it itself a complex number:

$$\Psi \in \mathbb{C}. \quad (2.7)$$

Since complex numbers cannot be measured the relation between the wavefunction and a measurable quantity must be somewhat indirect. Born's idea to identify probabilities, which are always positive real numbers, with the square of the norm of the wavefunction was very natural. If we write the wavefunction of our quantum system as Ψ , the probabilities for possible events are computed from $|\Psi|^2$. The mathematical framework required to express the laws of quantum mechanics consists of complex vector spaces. In any vector space we have objects called vectors that can be added together. In a complex vector space a vector multiplied by a complex number is still a vector. As we will see in our study of quantum mechanics it is many times useful to think of the wavefunction Ψ as a vector in some complex vector space.

3 Loss of Determinism

Maxwell's crowning achievement was the realization that his equations of electromagnetism allowed for the existence of propagating waves. In particular, in 1865 he conjectured that light was an electromagnetic wave, a propagating fluctuation of electric and magnetic fields. He was proven right in subsequent experiments. Towards the end of the nineteenth century physicists were convinced that light was a wave. The certainty, however, did not last too long. Experiments on blackbody radiation and on the photo-emission of electrons suggested that the behavior of light had to be more complicated than that of a simple wave. Max Planck and Albert Einstein were the most prominent contributors to the resolution of the puzzles raised by those experiments.

In order to explain the features of the photoelectric effect, Einstein postulated (1905) that in a light beam the energy comes in quanta – the beam is composed of packets of energy. Einstein essentially implied that light was made up of particles, each carrying a fixed amount of energy. He himself found this idea disturbing, convinced like most other contemporaries that, as Maxwell had shown, light was a wave. He anticipated that a physical entity, like light, that could behave both as a particle and as a wave could bring about the demise of classical physics and would require a completely new physical theory. He was in fact right. Though he never quite liked quantum mechanics, his ideas about particles of light, later given the name *photons*, helped construct this theory.

It took physicists until 1925 to accept that light could behave like a particle. The experiments of Arthur Compton (1923) eventually convinced most skeptics. Nowadays, particles of light, or photons, are routinely manipulated in laboratories around the world. Even if mysterious, we have grown accustomed to them. Each photon of visible light carries very little energy – a small laser pulse can

contain many billions of photons. Our eye, however, is a very good photon detector: in total darkness, we are able to see light when as little as ten photons hit upon our retina. When we say that light behaves like a particle we mean a quantum mechanical particle: a packet of energy and momentum that is not composed of smaller packets. We *do not* mean a classical point particle or Newtonian corpuscle, which is a zero-size object with definite position and velocity.

As it turns out, the energy of a photon depends only on the color of the light. As Einstein discovered the energy E and frequency ν for a photon are related by

$$E = h\nu \quad (3.1)$$

The frequency of a photon determines the wavelength λ of the light through the relation $\nu\lambda = c$, where c is the speed of light. All green photons, for example, have the same energy. To increase the energy in a light beam while keeping the same color, one simply needs more photons.

As we now explain, the existence of photons implies that Quantum Mechanics is not deterministic. By this we mean that the result of an experiment cannot be determined, as it would in classical physics, by the conditions that are under the control of the experimenter.

Consider a polarizer whose preferential direction is aligned along the \hat{x} direction, as shown in Figure 1. Light that is linearly polarized along the \hat{x} direction namely, light whose electric field points in this direction, goes through the polarizer. If the incident light polarization is orthogonal to the \hat{x} direction the light will not go through at all. Thus light linearly polarized in the \hat{y} direction will be totally absorbed by the polarizer. Now consider light polarized along a direction forming an angle α with the x -axis, as shown in Figure 2. What happens?

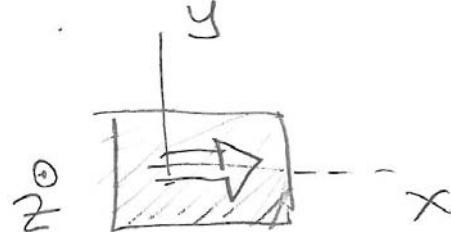


Figure 1: A polarizer that transmits light linearly polarized along the \hat{x} direction.

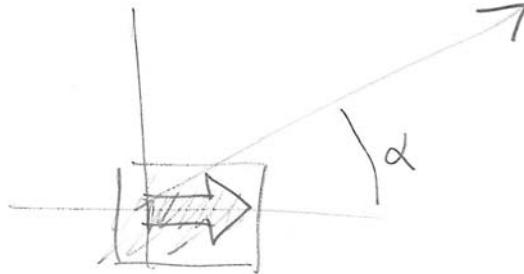


Figure 2: Light linearly polarized along the direction at an angle α hitting the polarizer.

Thinking of the light as a propagating wave, the incident electric field \mathbf{E}_α makes an angle α with the x -axis and therefore takes the form

$$\mathbf{E}_\alpha = E_0 \cos \alpha \hat{x} + E_0 \sin \alpha \hat{y}. \quad (3.2)$$

This is an electric field of magnitude E_0 . In here we are ignoring the time and space dependence of the wave; they are not relevant to our discussion. When this electric field hits the polarizer, the component along $\hat{\mathbf{x}}$ goes through and the component along $\hat{\mathbf{y}}$ is absorbed. Thus

$$\text{Beyond the polarizer: } \mathbf{E} = E_0 \cos \alpha \hat{\mathbf{x}}. \quad (3.3)$$

You probably recall that the energy in an electromagnetic wave is proportional to the square of the magnitude of the electric field. This means that the fraction of the beam's energy that goes through the polarizer is $(\cos \alpha)^2$. It is also well known that the light emerging from the polarizer has the *same frequency* as the incident light.

So far so good. But now, let us try to understand this result by thinking about the photons that make up the incident light. The premise here is that all photons in the incident beam are identical. Moreover the photons do not interact with each other. We could even imagine sending the whole energy of the incident light beam one photon at a time. Since all the light that emerges from the polarizer has the same frequency as the incident light, and thus the same frequency, we must conclude that each individual photon either goes through or is absorbed. If a fraction of a photon went through it would be a photon of lower energy and thus lower frequency, which is something that does not happen.

But now we have a problem. As we know from the wave analysis, roughly a fraction $(\cos \alpha)^2$ of the photons must go through, since that is the fraction of the energy that is transmitted. Consequently a fraction $1 - (\cos \alpha)^2$ of the photons must be absorbed. But if all the photons are identical, why is it that what happens to one photon does not happen to all of them?

The answer in quantum mechanics is that there is indeed a loss of determinism. No one can predict if a photon will go through or will get absorbed. The best anyone can do is to predict probabilities. In this case there would be a probability $(\cos \alpha)^2$ of going through and a probability $1 - (\cos \alpha)^2$ of failing to go through.

Two escape routes suggest themselves. Perhaps the polarizer is not really a homogeneous object and depending exactly on where the photon hits it either gets absorbed or goes through. Experiments show this is not the case. A more intriguing possibility was suggested by Einstein and others. A possible way out, they claimed, was the existence of *hidden variables*. The photons, while apparently identical, would have other *hidden* properties, not currently understood, that would determine with certainty which photon goes through and which photon gets absorbed. Hidden variable theories would seem to be untestable, but surprisingly they can be tested. Through the work of John Bell and others, physicists have devised clever experiments that rule out most versions of hidden variable theories. No one has figured out how to restore determinism to quantum mechanics. It seems to be an impossible task.

When we try to describe photons quantum mechanically we could use wavefunctions, or equivalently the language of states. A photon polarized along the $\hat{\mathbf{x}}$ direction is not represented using an electric field, but rather we just give a name for its *state*:

$$|\text{photon}; x\rangle. \quad (3.4)$$

We will learn the rules needed to manipulate such objects, but for the time being you could think of it like a vector in some space yet to be defined. Another state of a photon, or vector is

$$|\text{photon}; y\rangle, \quad (3.5)$$

representing a photon polarized along $\hat{\mathbf{y}}$. These states are the wavefunctions that represent the photon. We now claim that the photons in the beam that is polarized along the direction α are in a state $|\text{photon}; \alpha\rangle$ that can be written as a superposition of the above two states:

$$|\text{photon}; \alpha\rangle = \cos \alpha |\text{photon}; x\rangle + \sin \alpha |\text{photon}; y\rangle. \quad (3.6)$$

This equation should be compared with (3.2). While there are some similarities –both are superpositions– one refers to electric fields and the other to “states” of a single photon. Any photon that emerges from the polarizer will necessarily be polarized in the $\hat{\mathbf{x}}$ direction and therefore it will be in the state

$$\text{Beyond the polarizer: } |\text{photon}; x\rangle. \quad (3.7)$$

This can be compared with (3.3) which with the factor $\cos \alpha$ carries information about the amplitude of the wave. Here, for a single photon, there is no room for such a factor.

In the famous Fifth Solvay International Conference of 1927 the world’s most notable physicists gathered to discuss the newly formulated quantum theory. Seventeen out of the twenty nine attendees were or became Nobel Prize winners. Einstein, unhappy with the uncertainty in quantum mechanics stated the nowadays famous quote: “God does not play dice”, to which Niels Bohr is said to have answered: “Einstein, stop telling God what to do.” Bohr was willing to accept the loss of determinism, Einstein was not.

4 Quantum Superpositions

We have already discussed the concept of linearity; the idea that the sum of two solutions representing physical realities represents a new, allowed, physical reality. This superposition of solutions has a straightforward meaning in classical physics. In the case of electromagnetism, for example, if we have two solutions, each with its own electric and magnetic field, the “sum” solution is simply understood: its electric field is the sum of the electric fields of the two solutions and its magnetic field is the sum of the magnetic fields of the two solutions. In quantum mechanics, as we have explained, linearity holds. The interpretation of a superposition, however, is very surprising.

One interesting example is provided by a Mach-Zehnder interferometer; an arrangement of beam splitters, mirrors, and detectors used by Ernst Mach and Ludwig Zehnder in the 1890’s to study interference between two beams of light.

A beam splitter, as its name indicates, splits an incident beam into two beams, one that is reflected from the splitter and one that goes through the splitter. Our beam-splitters will be balanced: they split a given beam into two beams of equal intensity (Figure 3). The light that bounces off is called the reflected beam, the light that goes through is called the transmitted beam. The incident beam can hit the beam splitter from the top or from the bottom.

The Mach-Zehnder configuration, shown in Figure 4, has a left beam splitter (BS1) and a right beam splitter (BS2). In between we have the two mirrors, M1 on the top and M2 on the bottom. An incoming beam from the left is split by BS1 into two beams, each of which hits a mirror and is then sent into BS2. At BS2 the beams are recombined and sent into two outgoing beams that go into photon detectors D0 and D1.

It is relatively simple to arrange the beam-splitters so that the incoming beam, upon splitting at BS1 and recombination at BS2 emerges in the top beam which goes into D0. In this arrangement no light at all goes into D1. This requires a precise interference effect at BS2. Note that we have

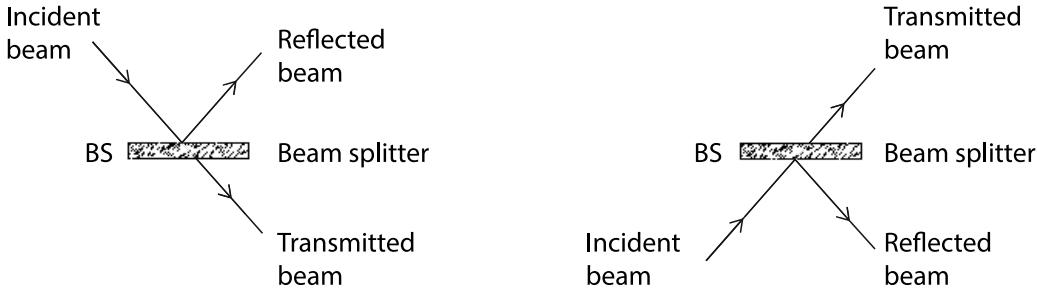


Figure 3: An incident beam hitting a beam-splitter results in a reflected beam and a transmitted beam. Left: incident beam coming from the top. Right: incident beam coming from the bottom.

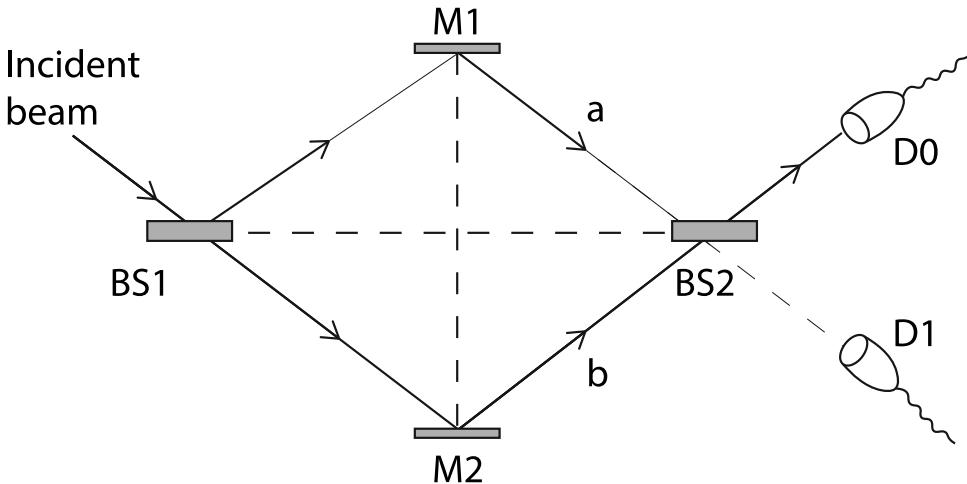


Figure 4: A Mach-Zehnder interferometer consists of two beam splitters BS1 and BS2, two mirrors M1 and M2, and two detectors D0 and D1. An incident beam will be split into two beams by BS1. One beam goes through the upper branch, which contains M1, the other beam goes through the lower branch, which contains M2. The beams on the two branches recombine at BS2 and are then sent into the detectors. The configuration is prepared to produce an interference so that all incident photons end at the detector D0, with none at D1.

two beams incident upon BS2; the top beam is called ‘a’ and the lower beam is called ‘b’. Two contributions go towards D0: the reflection of ‘a’ at BS2 and the transmission from ‘b’ at BS2. These two contributions interfere constructively to give a beam going into D0. Two contributions also go towards D1: the transmission from ‘a’ at BS2 and the reflection from ‘b’ at BS2. These two can indeed be arranged to interfere destructively to give no beam going into D1.

It is instructive to think of the incoming beam as a sequence of photons that we send into the interferometer, one photon at a time. This shows that, at the level of photons, the interference is not interference of one photon with another photon. Each photon must interfere with *itself* to give the result. Indeed interference between two photons is not possible: destructive interference, for example, would require that two photons end up giving no photon, which is impossible by energy conservation.

Therefore, each photon does the very strange thing of going through both branches of the interferometer! Each photon is in a superposition of two states: a state in which the photon is in the top beam or upper branch, added to a state in which the photon is in the bottom beam or lower branch. Thus the state of the photon in the interferometer is a funny state in which the photon seems to be doing two incompatible things at the same time.

Equation (3.6) is another example of a quantum superposition. The photon state has a component along an x -polarized photon and a component along a y -polarized photon.

When we speak of a wavefunction, we also sometimes call it a state, because the wavefunction specifies the “state” of our quantum system. We also sometimes refer to states as vectors. A quantum state may not be a vector like the familiar vectors in three-dimensional space but it is a vector nonetheless because it makes sense to add states and to multiply states by numbers. Just like vectors can be added, linearity guarantees that adding wavefunctions or states is a sensible thing to do. Just like any vector can be written as a sum of other vectors in many different ways, we will do the same with our states. By writing our physical state as sums of other states we can learn about the properties of our state.

Consider now two states $|A\rangle$ and $|B\rangle$. Assume, in addition, that when measuring some property Q in the state $|A\rangle$ the answer is always a , and when measuring the same property Q in the state $|B\rangle$ the answer is always b . Suppose now that our physical state $|\Psi\rangle$ is the superposition

$$|\Psi\rangle = \alpha|A\rangle + \beta|B\rangle, \quad \alpha, \beta \in \mathbb{C}. \quad (4.1)$$

What happens now if we measure property Q in the system described by the state $|\Psi\rangle$? It may seem reasonable that one gets some intermediate value between a and b , but this is not what happens. A measurement of Q will yield either a or b . There is no certain answer, classical determinism is lost, but the answer is always one of these two values and not an intermediate one. The coefficients α and β in the above superposition affect the probabilities with which we may obtain the two possible values. In fact, the probabilities to obtain a or b

$$\text{Probability}(a) \sim |\alpha|^2, \quad \text{Probability}(b) \sim |\beta|^2. \quad (4.2)$$

Since the only two possibilities are to measure a or b , the actual probabilities must sum to one and therefore they are given by

$$\text{Probability}(a) = \frac{|\alpha|^2}{|\alpha|^2 + |\beta|^2}, \quad \text{Probability}(b) = \frac{|\beta|^2}{|\alpha|^2 + |\beta|^2}. \quad (4.3)$$

If we obtain the value a , immediate repeated measurements would still give a , so the state after the measurement must be $|A\rangle$. The same happens for b , so we have

$$\begin{aligned} \text{After measuring } a \text{ the state becomes } |\Psi\rangle &= |A\rangle, \\ \text{After measuring } b \text{ the state becomes } |\Psi\rangle &= |B\rangle. \end{aligned} \quad (4.4)$$

In quantum mechanics one makes the following assumption: *Superposing a state with itself doesn't change the physics*, nor does it change the state in a non-trivial way. Since superimposing a state with itself simply changes the overall number multiplying it, we have that Ψ and $\alpha\Psi$ represent the same physics for any complex number α different from zero. Thus, letting \cong represent physical equivalence

$$|A\rangle \cong 2|A\rangle \cong i|A\rangle \cong -|A\rangle. \quad (4.5)$$

This assumption is necessary to verify that the polarization of a photon state has the expected number of degrees of freedom. The polarization of a plane wave, as one studies in electromagnetism, is described by two real numbers. For this consider an elliptically polarized wave, as shown in Figure 5. At any given point, the electric field vector traces an ellipse whose shape is encoded by the ratio a/b of

the semi-major axes (the first real parameter) and a tilt encoded by the angle θ (the second real parameter). Consider for this a general photon state formed by superposition of the two independent polarization states $|\text{photon}; x\rangle$ and $|\text{photon}; y\rangle$:

$$\alpha|\text{photon}; x\rangle + \beta|\text{photon}; y\rangle, \quad \alpha, \beta \in \mathbb{C}. \quad (4.6)$$

At first sight it looks as if we have two complex parameters α and β , or equivalently, four real parameters. But since the overall factor does not matter we can multiply this state by $1/\alpha$ to get the equivalent state that encodes all the physics

$$|\text{photon}; x\rangle + \frac{\beta}{\alpha}|\text{photon}; y\rangle, \quad (4.7)$$

showing that we really have one complex parameter, the ratio β/α . This is equivalent to two real parameters, as expected.

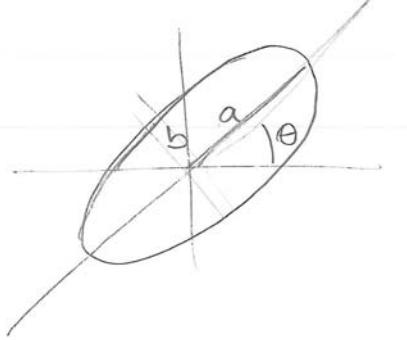


Figure 5: Parameters that define an elliptically polarized state.

Let us do a further example of superposition using electrons. Electrons are particles with spin. Classically, we imagine them as tiny balls spinning around an axis that goes through the particle itself. Once an axis is fixed, the electron has two and only two options: its rotation may be clockwise or counterclockwise about the axis, but in both cases it spins at the same fixed rate. These opposite ways of spinning are called *spin up* and *spin down* along the axis (see Figure 6). The up and down refer to the direction of the angular momentum associated with the rotation, and it is indicated by an arrow. According to quantum mechanics, and as verified by multiple experiments, the same possibilities, up or down, arise *whatever* axis we use to measure the spin of the electron.

Physicists usually set up coordinate systems in space by choosing three orthogonal directions, the directions of the x , y , and z axes. Let us choose to describe our spinning electrons using the z axis. One possible state of an electron is to be spin up along the z axis. Such a state is described as $|\uparrow; z\rangle$, with an arrow pointing up, and the label z indicating that the spin arrow points along the increasing z direction. Another possible state of an electron is spin down along the z axis. Such a state is described as $|\downarrow; z\rangle$, with an arrow pointing down, meaning this time that the spin points along the decreasing z direction. If these two are possible realities, so it would be the state $|\Psi\rangle$ representing the sum

$$|\Psi\rangle = |\uparrow; z\rangle + |\downarrow; z\rangle.$$

The state $|\Psi\rangle$ is in a superposition of a spin up and a spin down state. What kind of physics does this sum $|\Psi\rangle$ represent? It represents a state in which a measurement of the spin along the z axis would result in two possible outcomes with equal probabilities: an electron with spin up or an electron with spin down. Since we can only speak of probabilities, any experiment must involve repetition until

probabilities can be determined. Suppose we had a large ensemble of such electrons, all of them in the above state $|\Psi\rangle$. As we measured their spin along z , one at a time, we would find about half of them spinning up along z and the other half spinning down along z . There is no way to predict which option will be realized as we measure each electron. It is not easy to imagine superposition, but one may try as follows. An electron in the above state is in a different kind of existence in which it is able to both be spinning up along z and spinning down along z simultaneously! It is in such a ghostly, eerie state, doing incompatible things simultaneously, until its spin is measured. Once measured, the electron must immediately choose one of the two options; we always find electrons either spinning up or spinning down.

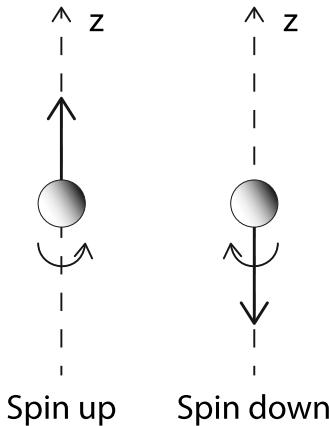


Figure 6: An electron with spin along the z axis. Left: the electron is said to have spin up along z . Right: the electron is said to have spin down along z . The up and down arrows represent the direction of the angular momentum associated with the spinning electron.

A critic of quantum mechanics could suggest a simpler explanation for the above observations. He or she would claim that the following simpler ensemble results in identical experimental results. In the critic's ensemble we have a large number of electrons with 50% of them in the state $|\uparrow; z\rangle$ and 50% of them in the state $|\downarrow; z\rangle$. He or she would then state, correctly, that such an ensemble would yield the same measurements of spins along z as the ensemble of those esoteric $|\Psi\rangle$ states. The new ensemble could provide a simpler explanation of the result without having to invoke quantum superpositions.

Quantum mechanics, however, allows for further experiments that can distinguish between the ensemble of our friendly critic and the ensemble of $|\Psi\rangle$ states. While it would take us too far afield to explain this, if we measured the spin of the electrons in the x direction, instead of z direction, the results would be *different* in the two ensembles. In the ensemble of our critic we would find 50% of the electrons up along x and 50% of the electrons down along x . In our ensemble of $|\Psi\rangle$ states, however, we would find a very simple result: all states pointing up along x . The critic's ensemble is not equivalent to our quantum mechanical ensemble. The critic is thus shown wrong in his or her attempt to show that quantum mechanical superpositions are not required.

5 Entanglement

When we consider superposition of states of *two* particles we can get the remarkable phenomenon called *quantum mechanical entanglement*. Entangled states of two particles are those in which we can't speak separately of the state of each particle. The particles are bound together in a common

state in which they are *entangled* with each other.

Let us consider two non-interacting particles. Particle 1 could be in any of the states

$$\{|u_1\rangle, |u_2\rangle, \dots\}, \quad (5.1)$$

while particle 2 could be in any of the states

$$\{|v_1\rangle, |v_2\rangle, \dots\} \quad (5.2)$$

It may seem reasonable to conclude that the state of the full system, including particle 1 and particle 2 would be specified by stating the state of particle 1 and the state of particle 2. If that would be the case the possible states would be written as

$$|u_i\rangle \otimes |v_j\rangle, \quad i, j \in \mathbb{N}, \quad (5.3)$$

for some specific choice of i and j that specify the state of particle one and particle two, respectively. Here we have used the symbol \otimes , which means *tensor* product, to combine the two states into a single state for the whole system. We will study \otimes later, but for the time being we can think of it as a kind of product that distributes over addition and obeys simple rules, as follows

$$\begin{aligned} (\alpha_1|u_1\rangle + \alpha_2|u_2\rangle) \otimes (\beta_1|v_1\rangle + \beta_2|v_2\rangle) = & \alpha_1\beta_1|u_1\rangle \otimes |v_1\rangle + \alpha_1\beta_2|u_1\rangle \otimes |v_2\rangle \\ & + \alpha_2\beta_1|u_2\rangle \otimes |v_1\rangle + \alpha_2\beta_2|u_2\rangle \otimes |v_2\rangle. \end{aligned} \quad (5.4)$$

The numbers can be moved across the \otimes but the order of the states must be preserved. The state on the left-hand side –expanded out on the right-hand side– is still of the type where we combine a state of the first particle ($\alpha_1|u_1\rangle + \alpha_2|u_2\rangle$) with a state of the second particle ($\beta_1|v_1\rangle + \beta_2|v_2\rangle$). Just like any one of the states listed in (5.3) this state is not entangled.

Using the states in (5.3), however, we can construct more intriguing superpositions. Consider the following one

$$|u_1\rangle \otimes |v_1\rangle + |u_2\rangle \otimes |v_2\rangle. \quad (5.5)$$

A state of two particles is said to be **entangled** if it cannot be written in the factorized form $(\dots) \otimes (\dots)$ which allows us to describe the state by simply stating the state of each particle. We can easily see that the state (5.5) cannot be factorized. If it could it would have to be with a product as indicated in (5.4). Clearly, involving states like $|u_3\rangle$ or $|v_3\rangle$ that do not appear in (5.5) would not help. To determine the constants $\alpha_1, \alpha_2, \beta_1, \beta_2$ we compare the right hand side of (5.4) with our state and conclude that we need

$$\alpha_1\beta_1 = 1, \quad \alpha_1\beta_2 = 0, \quad \alpha_2\beta_1 = 0, \quad \alpha_2\beta_2 = 1. \quad (5.6)$$

It is clear that there is no solution here. The second equation, for example, requires either α_1 or β_2 to be zero. Having $\alpha_1 = 0$ contradicts the first equation, and having $\beta_2 = 0$ contradicts the last equation. This confirms that the state (5.5) is indeed an entangled state. There is no way to describe the state by specifying a state for each of the particles.

Let us illustrate the above discussion using electrons and their spin states. Consider a state of two electrons denoted as $|\uparrow\rangle \otimes |\downarrow\rangle$. As the notation indicates, the first electron, described by the first arrow, is up along z while the second electron, described by the second arrow, is down along z (we omit the label z on the state for brevity). This is not an entangled state. Another possible state is

one where they are doing exactly the opposite: in $|\downarrow\rangle \otimes |\uparrow\rangle$ the first electron is down and the second is up. This second state is also not entangled. It now follows that by superposition we can consider the state

$$|\uparrow\rangle \otimes |\downarrow\rangle + |\downarrow\rangle \otimes |\uparrow\rangle. \quad (5.7)$$

This is an entangled state of the pair of electrons.

Exercise. Show that the above state cannot be factorized and thus is indeed entangled.

In the state (5.7) the first electron is up along z if the second electron is down along z (first term), or the first electron is down along z if the second electron is up along z (second term). There is a correlation between the spins of the two particles; they always point in opposite directions. Imagine that the two entangled electrons are very far away from each other: Alice has one electron of the pair on planet earth and Bob has the other electron on the moon. Nothing we know is connecting these particles but nevertheless the states of the electrons are linked. Measurements we do on the separate particles exhibit correlations. Suppose Alice measures the spin of the electron on earth. If she finds it up along z , it means that the first summand in the above superposition is realized, because in that summand the first particle is up. As discussed before, the state of the two particles immediately becomes that of the first summand. This means that the electron on the moon will *instantaneously* go into the spin down-along- z configuration, something that could be confirmed by Bob, who is sitting in the moon with that particle in his lab. This effect on Bob's electron happens before a message, carried with the speed of light, could reach the moon telling him that a measurement has been done by Alice on the earth particle and the result was spin up. Of course, experiments must be done with an ensemble that contains many pairs of particles, each pair in the same entangled state above. Half of the times the electron on earth will be found up, with the electron on the moon down and the other half of the times the electron on earth will be found down, with the electron on the moon up.

Our friendly critic could now say, correctly, that such correlations between the measurements of spins along z could have been produced by preparing a *conventional* ensemble in which 50% of the pairs are in the state $|\uparrow\rangle \otimes |\downarrow\rangle$ and the other 50% of the pairs are in the state $|\downarrow\rangle \otimes |\uparrow\rangle$. Such objections were dealt with conclusively in 1964 by John Bell, who showed that if Alice and Bob are able to measure spin in *three* arbitrary directions, the correlations predicted by the quantum entangled state are different from the classical correlations of *any* conceivable conventional ensemble. Quantum correlations in entangled states are very subtle and it takes sophisticated experiments to show they are not reproducible as classical correlations. Indeed, experiments with entangled states have confirmed the existence of quantum correlations. The kind of instantaneous action at a distance associated with measurements on well-separated entangled particles does not lead to paradoxes nor, as it may seem, to contradictions with the ideas of special relativity. You cannot use quantum mechanical entangled states to send information faster than the speed of light.

Sarah Geller transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare

<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

SPIN ONE-HALF, BRAS, KETS, AND OPERATORS

B. Zwiebach

September 17, 2013

Contents

1 The Stern-Gerlach Experiment	1
2 Spin one-half states and operators	7
3 Properties of Pauli matrices and index notation	12
4 Spin states in arbitrary direction	16

1 The Stern-Gerlach Experiment

In 1922, at the University of Frankfurt (Germany), Otto Stern and Walther Gerlach, did fundamental experiments in which beams of silver atoms were sent through inhomogeneous magnetic fields to observe their deflection. These experiments demonstrated that these atoms have quantized magnetic moments that can take two values. Although consistent with the idea that the electron had spin, this suggestion took a few more years to develop.

Pauli introduced a “two-valued” degree of freedom for electrons, without suggesting a physical interpretation. Kronig suggested in 1925 that it this degree of freedom originated from the self-rotation of the electron. This idea was severely criticized by Pauli, and Kronig did not publish it. In the same year Uhlenbeck and Goudsmit had a similar idea, and Ehrenfest encouraged them to publish it. They are presently credited with the discovery that the electron has an intrinsic spin with value “one-half”. Much of the mathematics of spin one-half was developed by Pauli himself in 1927. It took in fact until 1927 before it was realized that the Stern-Gerlach experiment did measure the magnetic moment of the electron.

A current on a closed loop induces a magnetic dipole moment. The magnetic moment vector $\vec{\mu}$ is proportional to the current I on the loop and the area A of the loop:

$$\vec{\mu} = IA \vec{A}. \quad (1.1)$$

The vector area, for a planar loop is a vector normal to the loop and of length equal to the value of the area. The direction of the normal is determined from the direction of the current and the right-hand rule. The product μB of the magnetic moment times the magnetic field has units of energy, thus the units of μ are

$$[\mu] = \frac{\text{erg}}{\text{gauss}} \text{ or } \frac{\text{Joule}}{\text{Tesla}} \quad (1.2)$$

When we have a charge distribution spinning we get a magnetic moment and, if the distribution has mass, an angular momentum. The magnetic moment and the angular momentum are proportional to each other, and the constant of proportionality is universal. To see this consider rotating radius R ring of charge with uniform charge distribution and total charge Q . Assume the ring is rotating about an axis perpendicular to the plane of the ring and going through its center. Let the tangential velocity at the ring be v . The current at the loop is equal to the linear charge density λ times the velocity:

$$I = \lambda v = \frac{Q}{2\pi R} v. \quad (1.3)$$

It follows that the magnitude μ of the dipole moment of the loop is

$$\mu = IA = \frac{Q}{2\pi R} v \pi R^2 = \frac{Q}{2} Rv. \quad (1.4)$$

Let the mass of the ring be M . The magnitude L of the angular momentum of the ring is then $L = R(Mv)$. As a result

$$\mu = \frac{Q}{2M} RMv = \frac{Q}{2M} L, \quad (1.5)$$

leading to the notable ratio

$$\frac{\mu}{L} = \frac{Q}{2M}.$$

(1.6)

Note that the ratio does not depend on the radius of the ring, nor on its velocity. By superposition, any rotating distribution with uniform mass and charge density will have a ratio μ/L as above, with Q the total charge and M the total mass. The above is also written as

$$\mu = \frac{Q}{2M} L. \quad (1.7)$$

an classical electron going in a circular orbit around a nucleus will have both orbital angular momentum and a magnetic moment, related as above, with Q the electron charge and M the electron mass. In quantum mechanics the electron is not actually going in circles around the proton, but the right quantum analog exhibits both orbital angular momentum and magnetic moment.

We can ask if the electron can have an intrinsic μ , as if it were, a tiny spinning ball. Well, it has an intrinsic μ but it cannot really be viewed as a rotating little ball of charge (this was part of Pauli's objection to the original idea of spin). Moreover, we currently view the electron as an elementary particle with zero size, so the idea that it rotates is just not sensible. The classical relation, however, points to the correct result. Even if it has no size, the electron has an intrinsic spin S –intrinsic angular momentum. One could guess that

$$\mu = \frac{e}{2m_e} S ? \quad (1.8)$$

Since angular momentum and spin have the same units we write this as

$$\mu = \frac{e\hbar}{2m_e} \frac{S}{\hbar} ? \quad (1.9)$$

This is not exactly right. For electrons the magnetic moment is twice as large as the above relation suggests. One uses a constant “ g -factor” to describe this

$$\mu = g \frac{e\hbar}{2m_e} \frac{S}{\hbar}, \quad g = 2 \text{ for an electron.} \quad (1.10)$$

This factor of two is in fact predicted by the Dirac equation for the electron, and has been verified experimentally. To describe the above more briefly, one introduces the canonical value μ_B of the dipole moment called the Bohr-magneton:

$$\mu_B \equiv \frac{e\hbar}{2m_e} = 9.27 \times 10^{-24} \frac{\text{J}}{\text{Tesla}}. \quad (1.11)$$

With this formula we get

$$\mu = g \mu_B \frac{S}{\hbar}, \quad g = 2 \text{ for an electron.} \quad (1.12)$$

Both the magnetic moment and the angular momentum point in the same direction if the charge is positive. For the electron we thus get

$$\vec{\mu} = -g \mu_B \frac{\vec{S}}{\hbar}, \quad g = 2. \quad (1.13)$$

Another feature of magnetic dipoles is needed for our discussion. A dipole placed in a non-uniform magnetic field will experience a force. An illustration is given in Figure 1 below, where to the left we show a current ring whose associated dipole moment $\vec{\mu}$ points upward. The magnetic field lines diverge as we move up, so the magnetic field is stronger as we move down. This dipole will experience a force pointing down, as can be deduced by basic considerations. On a small piece of wire the force $d\vec{F}$ is proportional to $\vec{I} \times \vec{B}$. The vectors $d\vec{F}$ are sketched in the right part of the figure. Their horizontal components cancel out, but the result is a net force downwards.

In general the equation for the force on a dipole $\vec{\mu}$ in a magnetic field \vec{B} is given by

$$\vec{F} = \nabla(\vec{\mu} \cdot \vec{B}). \quad (1.14)$$

Note that the force points in the direction for which $\vec{\mu} \cdot \vec{B}$ increases the fastest. Given that in our situation $\vec{\mu}$ and \vec{B} are parallel, this direction is the direction in which the magnitude of \vec{B} increases the fastest.

The Stern-Gerlach experiment uses atoms of silver. Silver atoms have 47 electrons. Forty-six of them fill completely the $n = 1, 2, 3$, and 4 levels. The last electron is an $n = 5$ electron with zero orbital angular momentum (a $5s$ state). The only possible angular momentum is the intrinsic angular momentum of the last electron. Thus a magnetic dipole moment is also that of the last electron (the nucleus has much smaller dipole moment and can be ignored). The silver is vaporized in an oven and with a help of a collimating slit a narrow beam of silver atoms is send down to a magnet configuration.

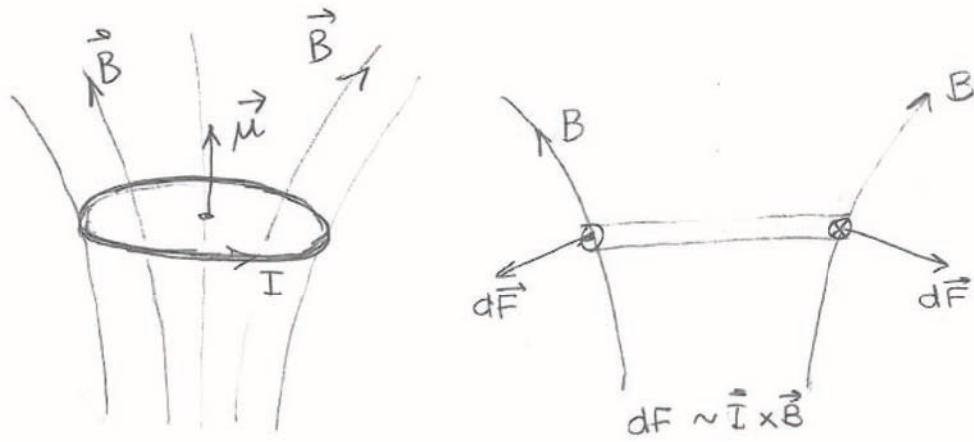


Figure 1: A magnetic dipole in a non-uniform magnetic field will experience a force. The force points in the direction for which $\vec{\mu} \cdot \vec{B}$ grows the fastest. In this case the force is downward.

In the situation described by Figure 2 the magnetic field points mostly in the positive z direction, and the gradient is also in the positive z -direction. As a result, the above equation gives

$$\vec{F} \simeq \nabla(\mu_z B_z) = \mu_z \nabla B_z \simeq \mu_z \frac{\partial B_z}{\partial z} \vec{e}_z, \quad (1.15)$$

and the atoms experience a force in the z -direction proportional to the z -component of their magnetic moment. Undeflected atoms would hit the detector screen at the point P . Atoms with positive μ_z should be deflected upwards and atoms with negative μ_z should be deflected downwards.

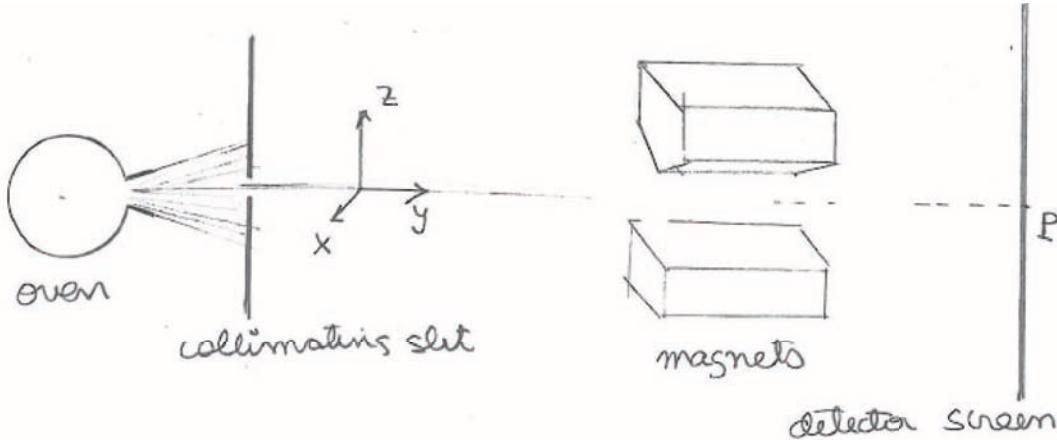


Figure 2: A sketch of the Stern-Gerlach apparatus. An oven and a collimating slit produces a narrow beam of silver atoms. The beam goes through a region with a strong magnetic field and a strong gradient, both in the z -direction. A screen, to the right, acts as a detector.

The oven source produces atoms with magnetic moments pointing in random directions and thus

the expectation was that the z-component of the magnetic moment would define a smooth probability distribution leading to a detection that would be roughly like the one indicated on the left side of Figure 3. Surprisingly, the observed result was two separate peaks as if all atoms had either a fixed positive μ_z or a fixed negative μ_z . This is shown on the right side of the figure. The fact that the peaks are spatially separated led to the original cumbersome name of “space quantization.” The Stern Gerlach experiment demonstrates the quantization of the dipole moment, and by theoretical inference from (1.13), the quantization of the spin (or intrinsic) angular momentum.

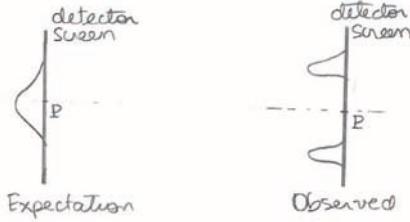


Figure 3: Left: the pattern on the detector screen that would be expected from classical physics. Right: the observed pattern, showing two separated peaks corresponding to up and down magnetic moments.

It follows from (1.13) that

$$\mu_z = -2\mu_B \frac{S_z}{\hbar}. \quad (1.16)$$

The deflections calculated using the details of the magnetic field configuration are consistent with

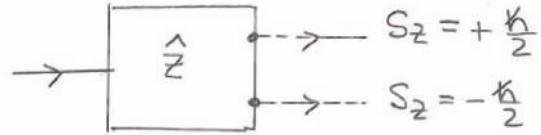
$$S_z = \pm \frac{\hbar}{2}, \text{ or } \frac{S_z}{\hbar} = \pm \frac{1}{2}. \quad (1.17)$$

A particle with such possible values of S_z/\hbar is called a spin one-half particle. The magnitude of the magnetic moments is one Bohr magneton.

With the magnetic field and its gradient along the z-direction, the Stern-Gerlach apparatus measures the component of the spin \vec{S} in the z direction. To streamline our pictures we will denote such apparatus as a box with a \hat{z} label, as in Figure 4. The box lets the input beam come in from the left and lets out two beams from the right side. If we placed a detector to the right, the top beam would be identified as having atoms with $S_z = \hbar/2$ and the bottom having atoms with $S_z = -\hbar/2$.¹

Let us now consider thought experiments in which we put a few SG apparatus in series. In the first configuration, shown at the top of Figure 5, the first box is a \hat{z} SG machine, where we block the $S_z = -\hbar/2$ output beam and let only the $S_z = \hbar/2$ beam go into the next machine. This machine acts as a filter. The second SG apparatus is also a \hat{z} machine. Since all ingoing particles have $S_z = \hbar/2$ the second machine lets those out the top output and nothing comes out in the bottom output. The quantum mechanical lesson here is that $S_z = \hbar/2$ states have no component or amplitude along $S_z = -\hbar/2$. These are said to be orthogonal states.

¹In the quantum mechanical view of the experiment, a single atom can be in both beams, with different amplitudes. Only the act of measurement, which corresponds to the act of placing the detector screen, forces the atom to decide in which beam it is.



SG machine measuring S_z

Figure 4: Left: A schematic representation of the SG apparatus, minus the screen.

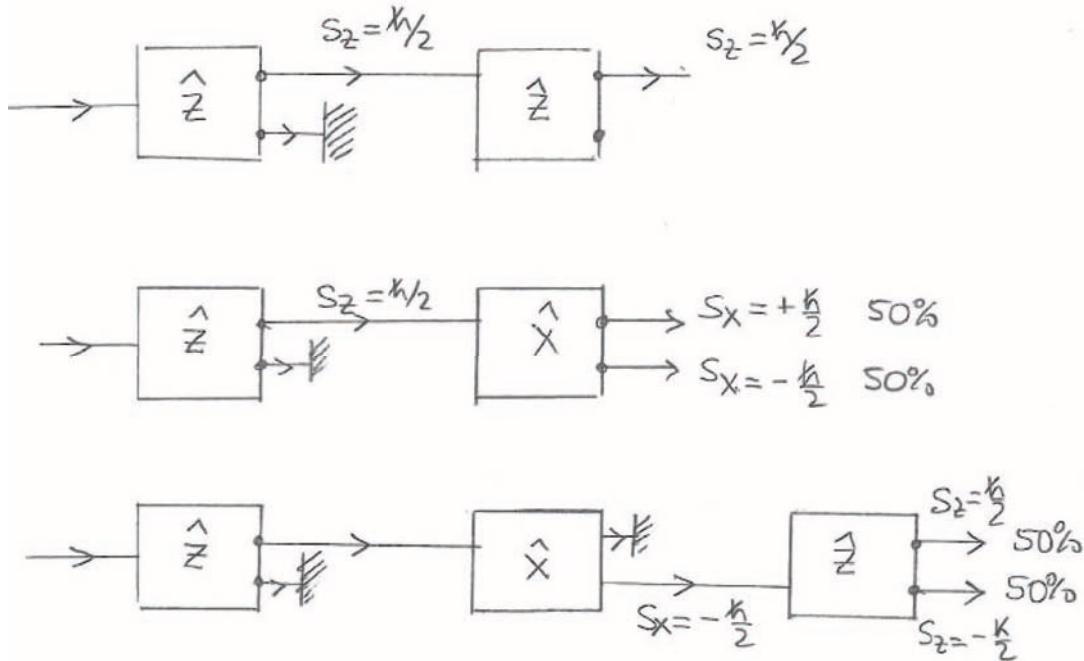


Figure 5: Left: Three configurations of SG boxes.

The second configuration in the figure shows the outgoing $S_z = \hbar/2$ beam from the first machine going into an \hat{x} -machine. The outputs of this machine are –in analogy to the \hat{z} machine– $S_x = \hbar/2$ and $S_x = -\hbar/2$. Classically an object with angular momentum along the z axis has no component of angular momentum along the x axis, these are orthogonal directions. But the result of the experiment indicates that quantum mechanically this is not true for spins. About half of the $S_z = \hbar/2$ atoms exit through the top $S_x = \hbar/2$ output, and the other half exit through the bottom $S_x = -\hbar/2$ output. Quantum mechanically, a state with a definite value of S_z has an amplitude along the state $S_x = \hbar/2$ as well as an amplitude along the state $S_x = -\hbar/2$.

In the third and bottom configuration the $S_z = \hbar/2$ beam from the first machine goes into the \hat{x} machine and the top output is blocked so that we only have an $S_x = -\hbar/2$ output. That beam is

fed into a \hat{z} type machine. One could speculate that the beam entering the third machine has both $S_x = -\hbar/2$ and $S_z = \hbar/2$, as it is composed of silver atoms that made it through both machines. If that were the case the third machine would let all atoms out the top output. This speculation is falsified by the result. There is no memory of the first filter: the particles out of the second machine do not anymore have $S_z = \hbar/2$. We find half of the particles make it out of the third machine with $S_z = \hbar/2$ and the other half with $S_z = -\hbar/2$. In the following section we discuss a mathematical framework consistent with the results of the above thought experiments.

2 Spin one-half states and operators

The SG experiment suggests that the spin states of the electron can be described using two basis vectors (or kets):

$$|z;+\rangle \quad \text{and} \quad |z;-\rangle. \quad (2.1)$$

The first corresponds to an electron with $S_z = \frac{\hbar}{2}$. The z label indicates the component of the spin, and the + the fact that the component of spin is positive. This state is also called ‘spin up’ along z . The second state corresponds to an electron with $S_z = -\frac{\hbar}{2}$, that is a ‘spin down’ along z . Mathematically, we have an operator \hat{S}_z for which the above states are eigenstates, with opposite eigenvalues:

$$\begin{aligned} \hat{S}_z |z;+\rangle &= +\frac{\hbar}{2} |z;+\rangle \\ \hat{S}_z |z;-\rangle &= -\frac{\hbar}{2} |z;-\rangle. \end{aligned} \quad (2.2)$$

If we have two basis states, then the state space of electron spin is a two-dimensional *complex vector space*. Each vector in this vector space represents a possible state of the electron spin. We are not discussing other degrees of freedom of the electron, such as its position, momentum, or energy. The general vector in the two-dimensional space is an arbitrary linear combination of the basis states and thus takes the form

$$|\Psi\rangle = c_1 |z;+\rangle + c_2 |z;-\rangle, \quad \text{with } c_1, c_2 \in \mathbb{C} \quad (2.3)$$

It is customary to call the state $|z;+\rangle$ the *first* basis state and it denote by $|1\rangle$. The state $|z;-\rangle$ is called the *second* basis state and is denoted by $|2\rangle$. States are vectors in some vector space. In a two-dimensional vector space a vector is explicitly *represented* as a column vector with two components. The first basis vector is represented as $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and the second basis vector is represented as $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Thus we have the following names for states and their concrete representation as column vectors

$$\begin{aligned} |z:+\rangle &= |1\rangle \longleftrightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ |z:-\rangle &= |2\rangle \longleftrightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned} \quad (2.4)$$

Using these options the state in (2.3) takes the possible forms

$$|\Psi\rangle = c_1|z;+\rangle + c_2|z;-\rangle = c_1|1\rangle + c_2|2\rangle \longleftrightarrow c_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \quad (2.5)$$

As we mentioned before, the top experiment in Figure 5 suggests that we have an orthonormal basis. The state $|z;+\rangle$ entering the second machine must have zero overlap with $|z,-\rangle$ since no such down spins emerge. Moreover the overlap of $|z;+\rangle$ with itself must be one, as all states emerge from the second machine top output. We thus write

$$\langle z;-|z;+\rangle = 0, \quad \langle z;+|z;+\rangle = 1. \quad (2.6)$$

and similarly, we expect

$$\langle z;+|z;-\rangle = 0, \quad \langle z;-|z;-\rangle = 1. \quad (2.7)$$

Using the notation where the basis states are labeled as $|1\rangle$ and $|2\rangle$ we have the simpler form that summarizes the four equations above:

$$\langle i|j\rangle = \delta_{ij}, \quad i, j = 1, 2. \quad (2.8)$$

We have not yet made precise what we mean by the ‘bras’ so let us do so briefly. We define the basis ‘bras’ as the *row vectors* obtained by transposition and complex conjugation:

$$\langle 1| \longleftrightarrow (1, 0), \quad \langle 2| \longleftrightarrow (0, 1). \quad (2.9)$$

Given states $|\alpha\rangle$ and $|\beta\rangle$

$$\begin{aligned} |\alpha\rangle &= \alpha_1|1\rangle + \alpha_2|2\rangle \longleftrightarrow \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \\ |\beta\rangle &= \beta_1|1\rangle + \beta_2|2\rangle \longleftrightarrow \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \end{aligned} \quad (2.10)$$

we associate

$$\langle \alpha| \equiv \alpha_1^* \langle 1| + \alpha_2^* \langle 2| \longleftrightarrow (\alpha_1^*, \alpha_2^*) \quad (2.11)$$

and the ‘bra-ket’ inner product is defined as the ‘obvious’ matrix product of the row vector and column vector representatives:

$$\langle \alpha|\beta\rangle \equiv (\alpha_1^*, \alpha_2^*) \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \alpha_1^*\beta_1 + \alpha_2^*\beta_2. \quad (2.12)$$

Note that this definition is consistent with (2.8).

When we represent the states as two-component column vectors the operators that act on the states to give new states can be *represented* as two-by-two matrices. We can thus represent the operator \hat{S}_z as a 2×2 matrix which we claim takes the form

$$\hat{S}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(2.13)

To test this, it suffices to verify that the matrix \hat{S}_z acts on the column vectors that represent the basis states as expected from (2.2). Indeed,

$$\begin{aligned}\hat{S}_z|z;+\rangle &= +\frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = +\frac{\hbar}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = +\frac{\hbar}{2}|z;+\rangle \\ \hat{S}_z|z;-\rangle &= +\frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{\hbar}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{\hbar}{2}|z;-\rangle.\end{aligned}\tag{2.14}$$

In fact, the states $|1\rangle$ and $|2\rangle$, viewed as column vectors are the eigenstates of matrix \hat{S}_z .

There is nothing particular about the z axis. We could have started with a SG apparatus that measures spin along the x axis and we would have been led to an operator \hat{S}_x . Had we used the y axis we would have been led to the operator \hat{S}_y . Since spin represents angular momentum (albeit of intrinsic type), it is expected to have three components, just like orbital angular momentum has three components: \hat{L}_x , \hat{L}_y , and \hat{L}_z . These are all hermitian operators, written as products of coordinates and momenta in three-dimensional space. Writing $\hat{L}_x = \hat{L}_1$, $\hat{L}_y = \hat{L}_2$, and $\hat{L}_z = \hat{L}_3$, their commutation relations can be briefly stated as

$$[\hat{L}_i, \hat{L}_j] = i\hbar\epsilon_{ijk}\hat{L}_k\tag{2.15}$$

This is the famous algebra of angular momentum, repeated indices are summed over the values 1,2,3, and ϵ_{ijk} is the totally antisymmetric symbol with $\epsilon_{123} = +1$. Make sure that you understand this notation clearly, and can use it to see that it implies the relations

$$\begin{aligned}[\hat{L}_x, \hat{L}_y] &= i\hbar\hat{L}_z, \\ [\hat{L}_y, \hat{L}_z] &= i\hbar\hat{L}_x, \\ [\hat{L}_z, \hat{L}_x] &= i\hbar\hat{L}_y.\end{aligned}\tag{2.16}$$

While, for example, $\hat{L}_z = \hat{x}\hat{p}_y - \hat{y}\hat{p}_x$ is a hermitian operator written in terms of coordinates and momenta, we have no such construction for \hat{S}_z . The latter is a more abstract operator, it does not act on wavefunctions $\psi(\vec{x})$ but rather on the 2-component column vectors introduced above. The operator \hat{S}_z is just a two-by-two *hermitian*² matrix with constant entries! If spin is a quantum mechanical angular momentum, we must have that the triplet of operators \hat{S}_z , \hat{S}_x , and \hat{S}_y satisfy

$$\begin{aligned}[\hat{S}_x, \hat{S}_y] &= i\hbar\hat{S}_z, \\ [\hat{S}_y, \hat{S}_z] &= i\hbar\hat{S}_x, \\ [\hat{S}_z, \hat{S}_x] &= i\hbar\hat{S}_y,\end{aligned}\tag{2.17}$$

or, again using numerical subscripts for the components ($\hat{S}_1 = \hat{S}_x, \dots$) we must have

$$[\hat{S}_i, \hat{S}_j] = i\hbar\epsilon_{ijk}\hat{S}_k.\tag{2.18}$$

²Hermitian means that the matrix is preserved by taking the operations of transposition and complex conjugation.

We can now try to figure out how the matrices for \hat{S}_x and \hat{S}_y must look, given that we know the matrix for \hat{S}_z . We have a few constraints. First the matrices must be hermitian, just like the angular momentum operators are. Two-by-two hermitian matrices take the form

$$\begin{pmatrix} 2c & a - ib \\ a + ib & 2d \end{pmatrix}, \text{ with } a, b, c, d \in \mathbb{R} \quad (2.19)$$

Indeed, you can easily see that transposing and complex conjugating gives exactly the same matrix. Since the two-by-two identity matrix commutes with every matrix, we can subtract from the above matrix any multiple of the identity without any loss of generality. Subtracting the identity matrix times $(c + d)$ we find the still hermitian matrix

$$\begin{pmatrix} c - d & a - ib \\ a + ib & d - c \end{pmatrix}, \text{ with } a, b, c, d \in \mathbb{R} \quad (2.20)$$

Since we are on the lookout for \hat{S}_x and \hat{S}_y we can subtract a matrix proportional to \hat{S}_z . Since \hat{S}_z is diagonal with entries of same value but opposite signs, we can cancel the diagonal terms above and are left over with

$$\begin{pmatrix} 0 & a - ib \\ a + ib & 0 \end{pmatrix}, \text{ with } a, b \in \mathbb{R} \quad (2.21)$$

Thinking of the space of two-by-two hermitian matrices as a real vector space, the hermitian matrices given above can be associated to two basis “vectors” that are the matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad (2.22)$$

since multiplying the first by the real constant a and the second by the real constant b and adding gives us the matrix above. In fact, together with the identity matrix and the \hat{S}_z matrix, with the $\hbar/2$ deleted,

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (2.23)$$

we got the complete set of four two-by-two matrices that viewed as basis vectors in a real vector space, can be used to build the most general hermitian two-by-two matrix by using real linear combinations.

Back to our problem, we are supposed to find \hat{S}_x and \hat{S}_y among the matrices in (2.22). The overall scale of the matrices can be fixed by the constraint that their eigenvalues be $\pm\hbar/2$, just like they are for \hat{S}_z . Let us give the eigenvalues (denoted by λ) and the associated normalized eigenvectors for these two matrices. Short computations (can you do them?) give

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} : \quad \lambda = 1, \text{ for } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \lambda = -1, \text{ for } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad (2.24)$$

for the first matrix and

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} : \quad \lambda = 1, \text{ for } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \lambda = -1, \text{ for } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}, \quad (2.25)$$

for the second matrix. In case you are puzzled by the normalizations, note that a vector $\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ is normalized if $|c_1|^2 + |c_2|^2 = 1$. Since the eigenvalues of both matrices are ± 1 , we tentatively identify

$$\hat{S}_x = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{S}_y = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad (2.26)$$

which have, at least, the correct eigenvalues. But in fact, these also satisfy the commutation relations! Indeed, we check that, as desired,

$$\begin{aligned} [\hat{S}_x, \hat{S}_y] &= \frac{\hbar^2}{4} \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} - \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) \\ &= \frac{\hbar^2}{4} \left(\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} - \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \right) \\ &= \frac{\hbar^2}{4} \begin{pmatrix} 2i & 0 \\ 0 & -2i \end{pmatrix} = i\hbar \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = i\hbar \hat{S}_z. \end{aligned} \quad (2.27)$$

All in all we have

$$\boxed{\hat{S}_x = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{S}_y = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{S}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.} \quad (2.28)$$

Exercise. Verify that the above matrices satisfy the other two commutation relations in (2.17).

You could ask if we got the unique solution for \hat{S}_x and \hat{S}_y given the choice of \hat{S}_z ? The answer is no, but it is not our fault, as illustrated by the following check:

Exercise. Check that the set of commutation relations of the spin operators are in fact preserved when we replace $\hat{S}_x \rightarrow -\hat{S}_y$ and $\hat{S}_y \rightarrow \hat{S}_x$.

The solution we gave is the one conventionally used by all physicists. Any other solution is physically equivalent to the one we gave (as will be explained in more detail after we develop more results). The solution defines the **Pauli matrices** σ_i by writing

$$\hat{S}_i = \frac{\hbar}{2} \sigma_i. \quad (2.29)$$

We then have that the Pauli matrices are

$$\boxed{\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.} \quad (2.30)$$

Let us describe the eigenstates of \hat{S}_x , which given (2.26) can be read from (2.24):

$$\hat{S}_x |x; \pm\rangle = \pm |x; \pm\rangle. \quad (2.31)$$

with

$$\boxed{\begin{aligned} |x; +\rangle &= \frac{1}{\sqrt{2}} |z; +\rangle + \frac{1}{\sqrt{2}} |z; -\rangle \longleftrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \\ |x; -\rangle &= \frac{1}{\sqrt{2}} |z; +\rangle - \frac{1}{\sqrt{2}} |z; -\rangle \longleftrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \end{aligned}} \quad (2.32)$$

Note that these states are orthogonal to each other. The above equations can be inverted to find

$$\begin{aligned} |z;+\rangle &= \frac{1}{\sqrt{2}}|x;+\rangle + \frac{1}{\sqrt{2}}|x;-\rangle \\ |z;-\rangle &= \frac{1}{\sqrt{2}}|x;+\rangle - \frac{1}{\sqrt{2}}|x;-\rangle \end{aligned} \quad (2.33)$$

These relations are consistent with the second experiment shown in Figure 5. The state $|z;+\rangle$ entering the second, \hat{x} -type SG apparatus, has equal probability to be found in $|x;+\rangle$ as it has probability to be found in $|x;-\rangle$. This is reflected in the first of the above relations, since we have the amplitudes

$$\langle x;+|z;+\rangle = \frac{1}{\sqrt{2}}, \quad \langle x;-|z;+\rangle = \frac{1}{\sqrt{2}}. \quad (2.34)$$

These probabilities, being equal to the norm squared of the amplitudes, are $1/2$ in both cases. The relative minus sign on the second equation above is needed to make it orthogonal to the state on the first equation.

We can finally consider the eigenstates of \hat{S}_y axis. We have

$$\hat{S}_y|y;\pm\rangle = \pm\frac{\hbar}{2}|y;\pm\rangle. \quad (2.35)$$

and using (2.25) we read

$$\begin{aligned} |y;+\rangle &= \frac{1}{\sqrt{2}}|z;+\rangle + \frac{i}{\sqrt{2}}|z;-\rangle \longleftrightarrow \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}, \\ |y;-\rangle &= \frac{1}{\sqrt{2}}|z;+\rangle - \frac{i}{\sqrt{2}}|z;-\rangle \longleftrightarrow \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}. \end{aligned} \quad (2.36)$$

Note that this time the superposition of $|z;\pm\rangle$ states involves complex numbers (there would be no way to find y type states without them).

3 Properties of Pauli matrices and index notation

Since we know the commutation relations for the spin operators

$$[\hat{S}_i, \hat{S}_j] = i\hbar\epsilon_{ijk}\hat{S}_k, \quad (3.37)$$

and we have $S_i = \frac{\hbar}{2}\sigma_i$, it follows that

$$\frac{\hbar}{2}\frac{\hbar}{2}[\sigma_i, \sigma_j] = i\hbar\epsilon_{ijk}\frac{\hbar}{2}\sigma_k. \quad (3.38)$$

Cancelling the \hbar 's and some factors of two, we find

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k. \quad (3.39)$$

Another important property of the Pauli matrices is that they square to the identity matrix. This is best checked explicitly (do it!):

$$(\sigma_1)^2 = (\sigma_2)^2 = (\sigma_3)^2 = \mathbf{1}. \quad (3.40)$$

This property “explains” that the eigenvalues of each of the Pauli matrices could only be plus or minus one. Indeed, the eigenvalues of a matrix satisfy the algebraic equation that the matrix satisfies. Take for example a matrix M that satisfies the matrix equation

$$M^2 + \alpha M + \beta \mathbf{1} = 0 \quad (3.41)$$

Let v be an eigenvector of M with eigenvalue λ : $Mv = \lambda v$. Let the above equation act on v

$$M^2 v + \alpha M v + \beta \mathbf{1} v = 0 \rightarrow \lambda^2 v + \alpha \lambda v + \beta v = 0 \rightarrow (\lambda^2 + \alpha \lambda + \beta)v = 0, \quad (3.42)$$

and since $v \neq 0$ (by definition an eigenvector cannot be zero!) we conclude that $\lambda^2 + \alpha \lambda + \beta = 0$, as claimed. For the case of the Pauli matrices we have $(\sigma_i)^2 = 1$ and therefore the eigenvalues must satisfy $\lambda^2 = 1$. As a result, $\lambda = \pm 1$ are the only options.

We also note, by inspection, that the Pauli matrices have zero trace, namely, the sum of entries on the diagonal is zero:

$$\text{tr}(\sigma_i) = 0, \quad i = 1, 2, 3. \quad (3.43)$$

A fact from linear algebra is that the trace of a matrix is equal to the sum of its eigenvalues. So each Pauli matrix must have two eigenvalues that add up to zero. Since the eigenvalues can only be plus or minus one, we must have one of each. This shows that each of the Pauli matrices has a plus one and a minus one eigenvalue.

If you compute a commutator of Pauli matrices by hand you might notice a curious property. Take the commutator of σ_1 and σ_2 :

$$[\sigma_1, \sigma_2] = \sigma_1 \sigma_2 - \sigma_2 \sigma_1. \quad (3.44)$$

The two contributions on the right hand side give

$$\begin{aligned} \sigma_1 \sigma_2 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \\ \sigma_2 \sigma_1 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}. \end{aligned} \quad (3.45)$$

The second contribution is minus the first, so that both terms contribute equally to the commutator! In other words,

$$\sigma_1 \sigma_2 = -\sigma_2 \sigma_1. \quad (3.46)$$

This equation is taken to mean that σ_1 and σ_2 *anticommute*. Just like we define the commutator of two operators X, Y by $[X, Y] \equiv XY - YX$, we define the **anticommutator**, denoted by curly brackets, by

$$\text{Anticommutator: } \{X, Y\} \equiv XY + YX. \quad (3.47)$$

In this language we have checked that

$$\{\sigma_1, \sigma_2\} = 0, \quad (3.48)$$

and the property $\sigma_1^2 = \mathbf{1}$, for example, can be rewritten as

$$\{\sigma_1, \sigma_1\} = 2 \cdot \mathbf{1}. \quad (3.49)$$

In fact, as you can check (two cases to examine) that any two different Pauli matrices anticommute:

$$\{\sigma_i, \sigma_j\} = 0, \quad \text{for } i \neq j. \quad (3.50)$$

We can easily improve on this equation to make it work also when i is equal to j . We claim that

$$\boxed{\{\sigma_i, \sigma_j\} = 2\delta_{ij} \mathbf{1}.} \quad (3.51)$$

Indeed, when $i \neq j$ the right-hand side vanishes, as needed, and when i is equal to j , the right-hand side gives $2 \cdot \mathbf{1}$, also as needed in view of (3.49) and its analogs for the other Pauli matrices.

Both the commutator and anti-commutator identities for the Pauli matrices can be summarized in a single equation. This is possible because, for any two operators X, Y we have

$$X Y = \frac{1}{2} \{X, Y\} + \frac{1}{2} [X, Y], \quad (3.52)$$

as you should confirm by expansion. Applied to the product of two Pauli matrices and using our expressions for the commutator and anticommutator we get

$$\boxed{\sigma_i \sigma_j = \delta_{ij} \mathbf{1} + i \epsilon_{ijk} \sigma_k.} \quad (3.53)$$

This equation can be recast in vector notation. Denote by bold symbols three-component vectors, for example, $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$. Then the dot product

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3 = a_i b_i = a_i b_j \delta_{ij}. \quad (3.54)$$

Note the use of the sum convention: repeated indices are summed over. Moreover, note that $b_j \delta_{ij} = b_i$ (can you see why?). We also have that

$$\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2. \quad (3.55)$$

Cross products use the epsilon symbol. Make sure you understand why

$$(\mathbf{a} \times \mathbf{b})_k = a_i b_j \epsilon_{ijk}. \quad (3.56)$$

We can also have triplets of operators, or matrices. For the Pauli matrices we denote

$$\boldsymbol{\sigma} \equiv (\sigma_1, \sigma_2, \sigma_3). \quad (3.57)$$

We can construct a matrix by dot product of a vector \mathbf{a} with the ‘vector’ $\boldsymbol{\sigma}$. We define

$$\mathbf{a} \cdot \boldsymbol{\sigma} \equiv a_1 \sigma_1 + a_2 \sigma_2 + a_3 \sigma_3 = a_i \sigma_i. \quad (3.58)$$

Note that $\mathbf{a} \cdot \boldsymbol{\sigma}$ is just a single two-by-two matrix. Since the components of \mathbf{a} are numbers, and numbers commute with matrices, this dot product is commutative: $\mathbf{a} \cdot \boldsymbol{\sigma} = \boldsymbol{\sigma} \cdot \mathbf{a}$. We are now ready to rewrite (3.53). Multiply this equation by $a_i b_j$ to get

$$\begin{aligned} a_i \sigma_i b_j \sigma_j &= a_i b_j \delta_{ij} \mathbf{1} + i (a_i b_j \epsilon_{ijk}) \sigma_k \\ &= (\mathbf{a} \cdot \mathbf{b}) \mathbf{1} + i (\mathbf{a} \times \mathbf{b})_k \sigma_k, \end{aligned} \quad (3.59)$$

so that, finally, we get the matrix equation

$$(\mathbf{a} \cdot \boldsymbol{\sigma})(\mathbf{b} \cdot \boldsymbol{\sigma}) = (\mathbf{a} \cdot \mathbf{b}) \mathbf{1} + i (\mathbf{a} \times \mathbf{b}) \cdot \boldsymbol{\sigma}. \quad (3.60)$$

As a simple application we take $\mathbf{b} = \mathbf{a}$. We then have $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$ and $\mathbf{a} \times \mathbf{a} = 0$, so that the above equation gives

$$(\mathbf{a} \cdot \boldsymbol{\sigma})^2 = |\mathbf{a}|^2 \mathbf{1}. \quad (3.61)$$

When \mathbf{a} is a unit vector this becomes

$$(\mathbf{n} \cdot \boldsymbol{\sigma})^2 = \mathbf{1}, \quad \mathbf{n} \text{ a unit vector.} \quad (3.62)$$

The epsilon symbol satisfies useful identities. One can show that the product of two epsilons with one index contracted is a sum of products of Kronecker deltas:

$$\epsilon_{ijk} \epsilon_{ipq} = \delta_{jp} \delta_{kq} - \delta_{jq} \delta_{kp}. \quad (3.63)$$

Its contraction (setting $p = j$) is also useful:

$$\epsilon_{ijk} \epsilon_{ijq} = 2\delta_{kq}. \quad (3.64)$$

The first of these two allows one to prove the familiar vector identity

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}. \quad (3.65)$$

It will be useful later on to consider the dot and cross products of *operator* triplets. Given the operators $\mathbf{X} = (\hat{X}_1, \hat{X}_2, \hat{X}_3)$ and $\mathbf{Y} = (\hat{Y}_1, \hat{Y}_2, \hat{Y}_3)$ we define

$$\begin{aligned} \mathbf{X} \cdot \mathbf{Y} &\equiv \hat{X}_i \hat{Y}_i, \\ (\mathbf{X} \times \mathbf{Y})_i &\equiv \epsilon_{ijk} \hat{X}_j \hat{Y}_k. \end{aligned} \quad (3.66)$$

In these definitions the order of the operators on the right hand side is as in the left-hand side. This is important to keep track of, since the \hat{X}_i and \hat{Y}_j operators may not commute. The dot product of two operator triplets is not necessarily commutative, nor is the cross product necessarily antisymmetric.

4 Spin states in arbitrary direction

We consider here the description and analysis of spin states that point in arbitrary directions, as specified by a unit vector \mathbf{n} :

$$\mathbf{n} = (n_x, n_y, n_z) = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta). \quad (4.67)$$

Here θ and ϕ are the familiar polar and azimuthal angles. We view the spatial vector \mathbf{n} as a triplet of numbers. Just like we did for $\boldsymbol{\sigma}$, we can define \mathbf{S} as the triplet of operators

$$\mathbf{S} = (\hat{S}_x, \hat{S}_y, \hat{S}_z). \quad (4.68)$$

Note that, in fact,

$$\mathbf{S} = \frac{\hbar}{2} \boldsymbol{\sigma}. \quad (4.69)$$

We can use \mathbf{S} to obtain, by a *dot* product with \mathbf{n} a spin operator $\hat{S}_{\mathbf{n}}$ that has a simple interpretation:

$$\hat{S}_{\mathbf{n}} \equiv \mathbf{n} \cdot \mathbf{S} \equiv n_x \hat{S}_x + n_y \hat{S}_y + n_z \hat{S}_z = \frac{\hbar}{2} \mathbf{n} \cdot \boldsymbol{\sigma}. \quad (4.70)$$

Note that $\hat{S}_{\mathbf{n}}$ is just an operator, or a hermitian matrix. We view $\hat{S}_{\mathbf{n}}$ as the spin operator in the direction of the unit vector \mathbf{n} . To convince you that this makes sense note that, for example, when \mathbf{n} points along z , we have $(n_x, n_y, n_z) = (0, 0, 1)$ and $\hat{S}_{\mathbf{n}}$ becomes \hat{S}_z . The same holds, of course, for the x and y directions. Moreover, just like all the \hat{S}_i , the eigenvalues of $\hat{S}_{\mathbf{n}}$ are $\pm \hbar/2$. This is needed physically, since all directions are physically equivalent and those two values for spin must be the only allowed values for all directions. To see that this is true we first compute the square of the matrix $\hat{S}_{\mathbf{n}}$:

$$(\hat{S}_{\mathbf{n}})^2 = \left(\frac{\hbar}{2}\right)^2 (\mathbf{n} \cdot \boldsymbol{\sigma})^2 = \left(\frac{\hbar}{2}\right)^2, \quad (4.71)$$

using (3.62). Moreover, since the Pauli matrices are traceless so is $\hat{S}_{\mathbf{n}}$:

$$\text{tr}(\hat{S}_{\mathbf{n}}) = n_i \text{tr}(\hat{S}_i) = n_i \frac{\hbar}{2} \text{tr}(\sigma_i) = 0. \quad (4.72)$$

By the same argument we used for Pauli matrices, we conclude that the eigenvalues of $\hat{S}_{\mathbf{n}}$ are indeed $\pm \hbar/2$. For an arbitrary direction we can write the matrix $\hat{S}_{\mathbf{n}}$ explicitly:

$$\begin{aligned} \hat{S}_{\mathbf{n}} &= \frac{\hbar}{2} \left[n_x \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + n_y \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + n_z \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right] \\ &= \frac{\hbar}{2} \begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix} \\ &= \frac{\hbar}{2} \begin{pmatrix} \cos \theta & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta \end{pmatrix}. \end{aligned} \quad (4.73)$$

Since the eigenvalues of $\hat{S}_{\mathbf{n}}$ are $\pm \hbar/2$ the associated spin eigenstates, denoted as $|\mathbf{n}; \pm\rangle$, satisfy

$$\hat{S}_{\mathbf{n}} |\mathbf{n}; \pm\rangle = \pm \frac{\hbar}{2} |\mathbf{n}; \pm\rangle. \quad (4.74)$$

The states $|\mathbf{n};+\rangle$ and $|\mathbf{n};-\rangle$ represent, respectively, a spin state that points up along \mathbf{n} , and a spin state that points down along \mathbf{n} . We can also find the eigenvalues of the matrix $\hat{S}_\mathbf{n}$ by direct computation. The eigenvalues are the roots of the equation $\det(\hat{S}_\mathbf{n} - \lambda \mathbf{1}) = 0$:

$$\det \begin{pmatrix} \frac{\hbar}{2} \cos \theta - \lambda & \frac{\hbar}{2} \sin \theta e^{-i\phi} \\ \frac{\hbar}{2} \sin \theta e^{i\phi} & -\frac{\hbar}{2} \cos \theta - \lambda \end{pmatrix} = \lambda^2 - \frac{\hbar^2}{4} (\cos^2 \theta + \sin^2 \theta) = \lambda^2 - \frac{\hbar^2}{4} = 0. \quad (4.75)$$

The eigenvalues are thus $\lambda = \pm \hbar/2$, as claimed. To find the eigenvector v associated with the eigenvalue λ we must solve the linear equation $(\hat{S}_\mathbf{n} - \lambda \mathbf{1})v = 0$. We denote by $|\mathbf{n};+\rangle$ the eigenvector associated with the eigenvalue $\hbar/2$. For this eigenvector we write the ansatz

$$|\mathbf{n};+\rangle = c_1 |+\rangle + c_2 |-\rangle = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad (4.76)$$

where for notational simplicity $|\pm\rangle$ refer to the states $|z;\pm\rangle$. The eigenvector equation becomes $(\hat{S}_\mathbf{n} - \frac{\hbar}{2}\mathbf{1})|\mathbf{n};+\rangle = 0$ and explicitly reads

$$\frac{\hbar}{2} \begin{pmatrix} \cos \theta - 1 & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta - 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0. \quad (4.77)$$

Either equation gives the same relation between c_1 and c_2 . The top equation, for example gives

$$c_2 = e^{i\phi} \frac{1 - \cos \theta}{\sin \theta} c_1 = e^{i\phi} \frac{\sin \frac{\theta}{2}}{\cos \frac{\theta}{2}} c_1. \quad (4.78)$$

(Check that the second equation gives the same relation.) We want normalized states, and therefore

$$|c_1|^2 + |c_2|^2 = 1 \rightarrow |c_1|^2 \left[1 + \frac{\sin^2 \frac{\theta}{2}}{\cos^2 \frac{\theta}{2}} \right] = 1 \rightarrow |c_1|^2 = \cos^2 \frac{\theta}{2}. \quad (4.79)$$

Since the overall phase of the eigenstate is not observable we take the simplest option for c_1 :

$$c_1 = \cos \frac{\theta}{2}, \quad c_2 = \sin \frac{\theta}{2} \exp(i\phi), \quad (4.80)$$

that is

$$|\mathbf{n};+\rangle = \cos \frac{\theta}{2} |+\rangle + \sin \frac{\theta}{2} e^{i\phi} |-\rangle. \quad (4.81)$$

As a quick check we see that for $\theta = 0$, which corresponds to a unit vector $\mathbf{n} = \mathbf{e}_3$ along the plus z direction we get $|\mathbf{e}_3;+\rangle = |+\rangle$. Note that even though ϕ is ambiguous when $\theta = 0$, this does not affect our answer, since the term with ϕ dependence vanishes. In the same way one can obtain the normalized eigenstate corresponding to $-\hbar/2$. A simple phase choice gives

$$|\mathbf{n};-\rangle = \sin \frac{\theta}{2} |+\rangle - \cos \frac{\theta}{2} e^{i\phi} |-\rangle. \quad (4.82)$$

If we again consider the $\theta = 0$ direction, this time the ambiguity of ϕ remains in the term that contains the $|z;-\rangle$ state. It is convenient to multiply this state by the phase $-e^{-i\phi}$. Doing this, the pair of

eigenstates read³

$ \mathbf{n}; +\rangle = \cos \frac{\theta}{2} +\rangle + \sin \frac{\theta}{2} e^{i\phi} -\rangle,$	(4.83)
$ \mathbf{n}; -\rangle = -\sin \frac{\theta}{2} e^{-i\phi} +\rangle + \cos \frac{\theta}{2} -\rangle.$	

The vectors are normalized. Furthermore, they are orthogonal

$$\langle \mathbf{n}; - | \mathbf{n}; + \rangle = -\sin \frac{\theta}{2} e^{i\phi} \cos \frac{\theta}{2} + \cos \frac{\theta}{2} \sin \frac{\theta}{2} e^{i\phi} = 0. \quad (4.84)$$

Therefore, $|\mathbf{n}; +\rangle$ and $|\mathbf{n}; -\rangle$ are an orthonormal pair of states.

Let us verify that the $|\mathbf{n}; \pm\rangle$ reduce to the known results as \mathbf{n} points along the z , x , and y axes. Again, if $\mathbf{n} = (0, 0, 1) = \mathbf{e}_3$, we have $\theta = 0$, and hence

$$|\mathbf{e}_3; +\rangle = |+\rangle, \quad |\mathbf{e}_3; -\rangle = |-\rangle, \quad (4.85)$$

which are, as expected, the familiar eigenstates of \hat{S}_z . If we point along the x axis, $\mathbf{n} = (1, 0, 0) = \mathbf{e}_1$ which corresponds to $\theta = \pi/2$, $\phi = 0$. Hence

$$|\mathbf{e}_1; +\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle) = |x; +\rangle, \quad |\mathbf{e}_1; -\rangle = \frac{1}{\sqrt{2}}(-|+\rangle + |-\rangle) = -|x; -\rangle, \quad (4.86)$$

where we compared with (2.32). Note that the second state came out with an overall minus sign. Since overall phases (or signs) are physically irrelevant, this is the expected answer: we got the eigenvectors of \hat{S}_x . Finally, if $\mathbf{n} = (0, 1, 0) = \mathbf{e}_2$, we have $\theta = \pi/2$, $\phi = \pi/2$ and hence, with $e^{\pm i\phi} = \pm i$, we have

$$|\mathbf{e}_2; +\rangle = \frac{1}{\sqrt{2}}(|+\rangle + i|-\rangle) = |y; +\rangle, \quad |\mathbf{e}_2; -\rangle = \frac{1}{\sqrt{2}}(i|+\rangle + |-\rangle) = i\frac{1}{\sqrt{2}}(|+\rangle - i|-\rangle) = i|y; -\rangle \quad (4.87)$$

which are, up to a phase for the second one, the eigenvectors of \hat{S}_y .

³The formula (4.83) works nicely at the north pole ($\theta = 0$), but at the south pole ($\theta = \pi$) the ϕ ambiguity shows up again. If one works near the south pole multiplying the results in (4.83) by suitable phases will do the job. The fact that no formula works well unambiguously through the full the sphere is not an accident.

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Chapter 2: Experiments with photons

B. Zwiebach
February 9, 2016

Contents

1 Mach-Zehnder Interferometer	1
2 Elitzur-Vaidman Bombs	6

1 Mach-Zehnder Interferometer

We have discussed before the Mach-Zehnder interferometer, which we show again in Figure 1. It contains two beam-splitters BS1 and BS2 and two mirrors. Inside the interferometer we have two beams, one going over the upper branch and one going over the lower branch. This extends beyond BS2: the upper branch continues to D0 while the lower branch continues to D1.

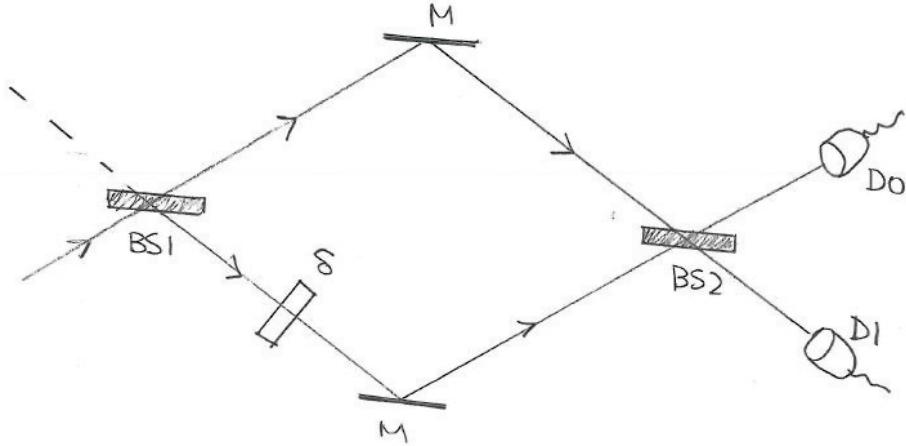


Figure 1: The Mach-Zehnder Interferometer

Vertical cuts in the above figure intersect the two beams and we can ask what is the probability to find a photon in each of the two beams at that cut. For this we need two probability *amplitudes*, or two complex numbers, whose norm-squared would give probabilities. We can encode this information in a two component vector as

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (1.1)$$

Here α is the probability amplitude to be in the upper beam and β the probability amplitude to be in the lower beam. Therefore, $|\alpha|^2$ would be the probability to find the photon in the upper beam and $|\beta|^2$ the probability to find the photon in the lower beam. Since the photon must be found in either one of the beams we must have

$$|\alpha|^2 + |\beta|^2 = 1. \quad (1.2)$$

Following this notation, we would have for the cases when the photon is definitely in one or the other beam:

$$\text{photon on upper beam: } \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{photon on bottom beam: } \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (1.3)$$

We can view the state (1.1) as a superposition of these two simpler states using the rules of vector addition and multiplication:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \beta \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (1.4)$$

In the interferometer shown in Figure 1 we included in the lower branch a ‘phase shifter’, a piece of



Figure 2: A phase shifter of phase factor $e^{i\delta}$. The amplitude gets multiplied by the phase.

material whose only effect is to multiply the probability amplitude by a fixed phase $e^{i\delta}$ with $\delta \in \mathbb{R}$. As shown in Figure 2, the probability amplitude α to the left of the device becomes $e^{i\delta}\alpha$ to the right of the device. Since the norm of a phase is one, the phase-shifter does not change the probability to find the photon. When the phase δ is equal to π the effect of the phase shifter is to change the sign of the wavefunction since $e^{i\pi} = -1$.

Let us now consider the effect of beam splitters in detail. If the incoming photon hits a beam-splitter from the top, we consider this photon to belong to the upper branch and represent it by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. If the incoming photon hits the beam-splitter from the bottom, we consider this photon to belong to the lower branch, and represent it by $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. We show the two cases in Figure 3. The effect of the beam splitter is to give an output wavefunction for each of the two cases:

$$\text{Left BS: } \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} s \\ t \end{pmatrix}, \quad \text{Right BS: } \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} u \\ v \end{pmatrix}. \quad (1.5)$$

As you can see from the diagram, for the photon hitting from above, s may be thought as a reflection amplitude and t as a transmission coefficient. Similarly, for the photon hitting from below, v may be thought as a reflection amplitude and u as a transmission coefficient. The four numbers s, t, u, v , by linearity, characterize completely the beam splitter. They can be used to predict the output given any incident photon, which may have amplitudes to hit both from above and from below. Indeed, an incident photon state $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ would give

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \alpha \begin{pmatrix} s \\ t \end{pmatrix} + \beta \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \alpha s + \beta u \\ \alpha t + \beta v \end{pmatrix} = \begin{pmatrix} s & u \\ t & v \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (1.6)$$

In summary, we see that the BS produces the following effect

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \rightarrow \begin{pmatrix} s & u \\ t & v \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (1.7)$$

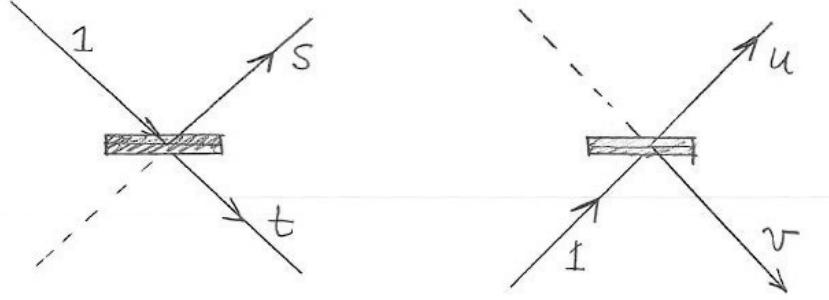


Figure 3: Left: A photon incident from the top; s and t are the reflected and transmitted amplitudes, respectively. Right: A photon incident from the bottom; v and u are the reflected and transmitted amplitudes, respectively.

We can represent the action of the beam splitter as matrix multiplication on the incoming wavefunction, with the two-by-two matrix

$$\begin{pmatrix} s & u \\ t & v \end{pmatrix}. \quad (1.8)$$

We must now figure out the constraints on s, t, u, v . Because probabilities must add up to one, equation (1.5) implies that

$$|s|^2 + |t|^2 = 1, \quad (1.9)$$

$$|u|^2 + |v|^2 = 1. \quad (1.10)$$

The kind of beam splitters we use are called balanced, which means that reflection and transmission probabilities are the same. So all four constants must have equal norm-squared:

$$|s|^2 = |t|^2 = |u|^2 = |v|^2 = \frac{1}{2}. \quad (1.11)$$

Let's try a guess for the values. Could we have

$$\begin{pmatrix} s & u \\ t & v \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} ? \quad (1.12)$$

This fails if acting on normalized wavefunctions (or column vectors) does not yield normalized wavefunctions. So we try with a couple of wavefunctions

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (1.13)$$

While the first example works out, the second does not, as $|1|^2 + |1|^2 = 2 \neq 1$. An easy fix is achieved by changing the sign of v :

$$\begin{pmatrix} s & u \\ t & v \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \quad (1.14)$$

Let's check that this matrix works in general. Thus acting on a state $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ with $|\alpha|^2 + |\beta|^2 = 1$ we find

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \alpha + \beta \\ \alpha - \beta \end{pmatrix}. \quad (1.15)$$

Indeed the resulting state is well normalized. The total probability is what we expect

$$\begin{aligned} \frac{1}{2}|\alpha + \beta|^2 + \frac{1}{2}|\alpha - \beta|^2 &= \frac{1}{2}(|\alpha|^2 + |\beta|^2 + \alpha\beta^* + \alpha^*\beta) + \frac{1}{2}(|\alpha|^2 + |\beta|^2 - \alpha\beta^* - \alpha^*\beta) \\ &= |\alpha|^2 + |\beta|^2 = 1. \end{aligned} \quad (1.16)$$

The minus sign in the bottom right entry of (1.14) means that a photon incident from below, as it is reflected, will have its amplitude changed by a sign or equivalently a phase shift by π (check this!). This effect, of course, is realized in practice. A typical beam splitter consists of a glass plate with a reflective dielectric coating on one side. The refractive index of the coating is chosen to be intermediate between that of glass and that of air. A reflection causes a phase shift only when light encounters a material of higher refractive index. This is the case in the transition of air to coating, but not in the transition from glass to coating. Thus the beam splitter represented by (1.14) would have its coating on the bottom side. Transmitted waves have no phase shift.

Another possibility for a beam splitter matrix is

$$\frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}, \quad (1.17)$$

which would be realized by a dielectric coating on the top side. You can quickly check that, like the previous matrix, its action also conserves probability. We will call the left beam-splitter BS1 and the right beam splitter BS2 and their respective matrices will be

$$\text{BS1 : } \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \text{BS2 : } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \quad (1.18)$$

The two beam splitters are combined to form the interferometer shown in Figure 4. If we now assume an input photon wavefunction $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ from the left, the output wavefunction that goes into the detectors is obtained by acting first with the BS1 matrix and then with the BS2 matrix:

$$\text{input : } \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \text{output: } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ -\alpha \end{pmatrix}. \quad (1.19)$$

With the help of this result, for any input photon state we can write immediately the output photon state that goes into the detectors.

If the input photon beam is $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, the output from the interferometer is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and therefore a photon will be detected at D0. This is shown in Figure 5. We can make a very simple table with the possible outcomes and their respective probabilities P :

Outcome	P
photon at D0	1
photon at D1	0

(1.20)

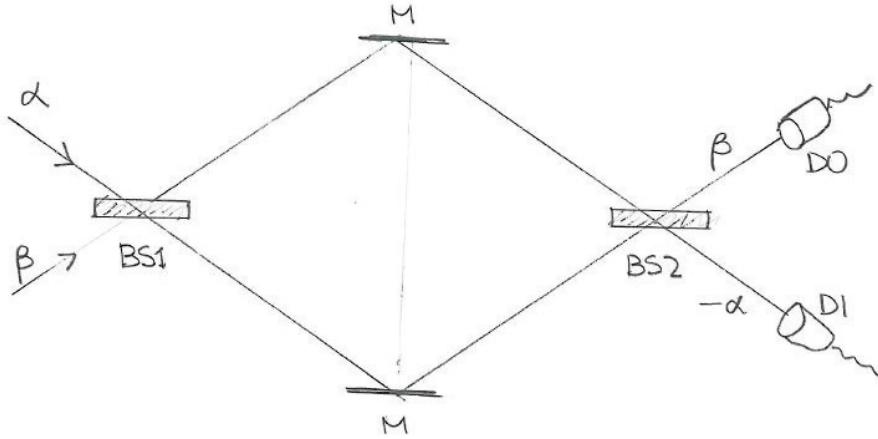


Figure 4: The Mach-Zehnder interferometer with input and output wavefunctions indicated.

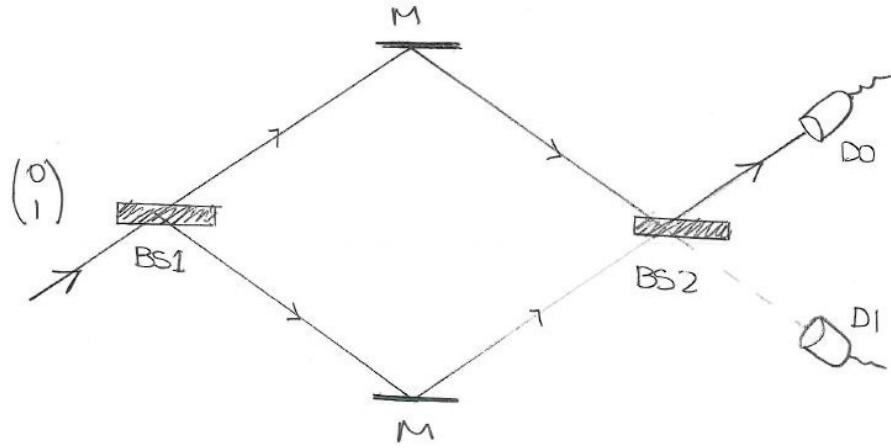


Figure 5: Incident photon from below will go into D0.

Now, block the lower path, as indicated in Figure 6. What happens then? It is best to track down things systematically. The input beam, acted by BS1 gives

$$\frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (1.21)$$

This is indicated in the figure, to the right of BS1. Then the lower branch is stopped, while the upper branch continues. The upper branch reaches BS2, and here the input is $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$, because nothing is coming from the lower branch. We therefore get an output

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}. \quad (1.22)$$

In this experiment there are three possible outcomes: the photon can be absorbed by the block, or

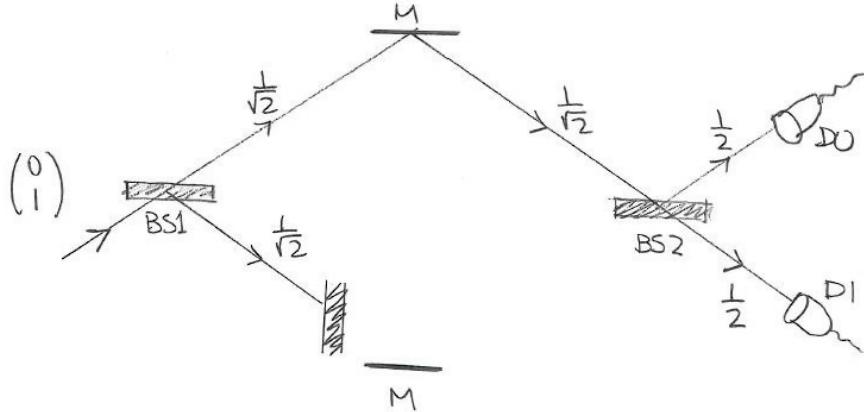


Figure 6: The probability to detect the photon at D1 can be changed by blocking one of the paths.

can go into any of the two detectors. As we see in the diagram, the probabilities are:

Outcome	P
photon at block	$\frac{1}{2}$
photon at D0	$\frac{1}{4}$
photon at D1	$\frac{1}{4}$

(1.23)

It is noteworthy that before blocking the lower path we could not get a photon to D1. The probability to reach D1 is now 1/4 and was increased by blocking a path.

2 Elitzur-Vaidman Bombs

To see that allowing the photon to reach D1 by blocking a path is very strange, we consider an imaginary situation proposed by physicists Avshalom Elitzur and Lev Vaidman, from Tel-Aviv University, in Israel. They imagined bombs with a special type of trigger: a photon detector. A narrow tube goes across each bomb and in the middle of the tube there is a photon detector. To detonate the bomb one sends a photon into the tube. The photon is then detected by the photon detector and the bomb explodes. If the photon detector is defective, however, the photon is not detected at all. It propagates freely through the tube and comes out of the bomb. The bomb does not explode.

Here is the situation we want to address. Suppose we have a number of Elitzur-Vaidman (EV) bombs, but we know that some of them have become defective. How could we tell if a bomb is operational without detonating it? Assume, for the sake of the problem, that we are unable to examine the detector without destroying the bomb.

We seem to be facing an impossible situation. If we send a photon into the detector tube and nothing happens we know the bomb is defective, but if the bomb is operational it would simply explode. It seems impossible to confirm that the photon detector in the bomb is working without testing it. Indeed, it is impossible in classical physics. It is not impossible in quantum mechanics, however. As we will see, we can perform what can be called an interaction-free measurement!

We now place an EV bomb on the lower path of the interferometer, with the detector tube properly aligned. Suppose we send in a photon as pictured. If the bomb is defective it is as if there is no detector, the lower branch of the interferometer is free and all the photons that we send in will end up in D0,

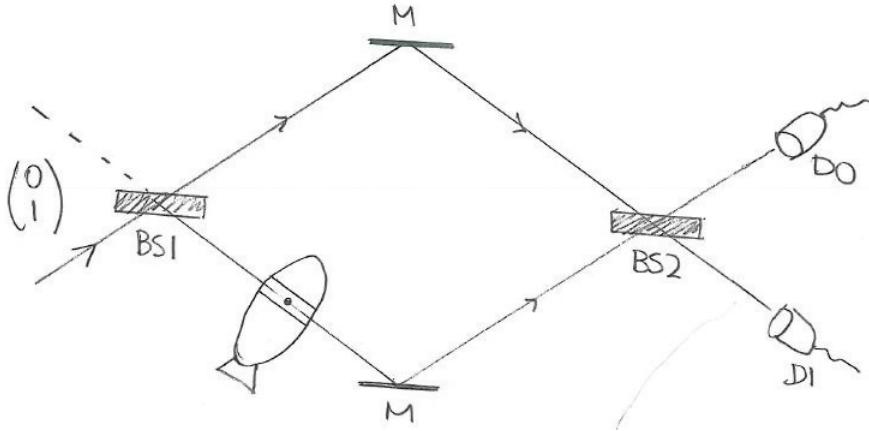


Figure 7: A Mach-Zehnder interferometer and an Elitzur-Vaidman bomb inserted on the lower branch, with the detector tube properly aligned. If the bomb is faulty all incident photons will end up at D0. If a photon ends up at D1 we know that the bomb is operational, even though the photon never went into the bomb detector!

just as they did in Figure 5.

Outcome	P	
photon at D0 no explosion	1	
photon at D1 no explosion	0	
bomb explodes	0	(2.24)

If the bomb is working, on the other hand, we have the situation we had in Figure 6, where we placed a block in the lower branch of the interferometer:

Outcome	P	
bomb explodes	$\frac{1}{2}$	
photon at D0 no explosion	$\frac{1}{4}$	
photon at D1 no explosion	$\frac{1}{4}$	(2.25)

Assume the bomb is working. Then 50% of the times the photon will hit it and it will explode, 25% of the time the photon will end in D0 and we can't tell if it is defective or not. But 25% of the time the photon will end in D1, and since this was impossible for a faulty bomb, we have learned that the bomb is operational! We have learned that even though the photon never made it through the bomb; it ended on D1. If you think about this you will surely realize it is extremely surprising and counterintuitive. But it is true, and experiments (without using bombs!) have confirmed that this kind of interaction-free measurement is indeed possible.

Sarah Geller transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare
<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

Lecture 6

B. Zwiebach
February 23, 2016

Contents

1 Normalization and time evolution	1
2 The Wavefunction as a Probability Amplitude	3
3 The Probability Current	5
4 Probability current in 3D and current conservation	7

1 Normalization and time evolution

The wavefunction $\Psi(x, t)$ that describes the quantum mechanics of a particle of mass m moving in a potential $V(x, t)$ satisfies the Schrödinger equation

$$i\hbar \frac{\partial \Psi(x, t)}{\partial t} = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x, t) \right) \Psi(x, t), \quad (1.1)$$

or more briefly

$$i\hbar \frac{\partial \Psi(x, t)}{\partial t} = \hat{H} \Psi(x, t). \quad (1.2)$$

The interpretation of the wavefunction arises by declaring that dP , defined by

$$dP = |\Psi(x, t)|^2 dx, \quad (1.3)$$

is the probability to find the particle in the interval dx centered on x at time t . It follows that the probabilities of finding the particle at all possible points must add up to one:

$$\int_{-\infty}^{\infty} \Psi^*(x, t) \Psi(x, t) dx = 1. \quad (1.4)$$

We will try to understand how this equation is compatible with the time evolution prescribed by the Schrödinger equation. But before that let us examine what kind of conditions are required from wavefunctions in order to satisfy (1.4).

Suppose the wavefunction has well-defined limits as $x \rightarrow \pm\infty$. If those limits are different from zero, the integral around infinity would produce an infinite result, which is inconsistent with the claim that the total integral is one. Therefore the limits should be zero:

$$\boxed{\lim_{x \rightarrow \pm\infty} \Psi(x, t) = 0.} \quad (1.5)$$

It is in principle possible to have a wavefunction that has no well-defined limit at infinity but is still square integrable. But such cases do not seem to appear in practice so we will assume that (1.5) holds. It would also be natural to assume that the spatial derivative of Ψ vanishes as $x \rightarrow \pm\infty$ but, as we will see soon, it suffices to assume that the limit of the spatial derivative of Ψ is bounded

$$\boxed{\lim_{x \rightarrow \pm\infty} \frac{\partial\Psi(x, t)}{\partial x} < \infty.} \quad (1.6)$$

We have emphasized before that the overall numerical factor multiplying the wavefunction is not physical. But equation (1.4) seems to be in conflict with this: if a given Ψ satisfies it, the presumed equivalent 2Ψ will not! To make precise sense of probabilities it is *convenient* to work with normalized wavefunctions, but it is not necessary, as we show now. Since time plays no role in the argument, so assume in all that follows that the equations refer to some time t_0 arbitrary but fixed. Suppose you have a wavefunction such that

$$\int dx |\Psi|^2 = \mathcal{N} \neq 1. \quad (1.7)$$

Then I claim that the probability dP to find the particle in the interval dx about x is given by

$$dP = \frac{1}{\mathcal{N}} |\Psi|^2 dx. \quad (1.8)$$

This is consistent because

$$\int dP = \frac{1}{\mathcal{N}} \int dx |\Psi|^2 = \frac{1}{\mathcal{N}} \cdot \mathcal{N} = 1. \quad (1.9)$$

Note that dP is not changed when Ψ is multiplied by any number. Thus, this picture makes it clear that the overall scale of Ψ contains no physics. As long as the integral $\int |\Psi|^2 dx < \infty$ the wavefunction is said to be **normalizable, or square-integrable**. By adjusting the overall coefficient of Ψ we can then make it **normalized**. Indeed, again assuming (1.7) the new wavefunction Ψ' defined by

$$\Psi' = \frac{1}{\sqrt{\mathcal{N}}} \Psi, \quad (1.10)$$

is properly normalized. Indeed

$$\int dx |\Psi'|^2 = \frac{1}{\mathcal{N}} \int |\Psi|^2 dx = 1. \quad (1.11)$$

We sometimes work with wavefunctions for which the integral (1.4) is infinite. Such wavefunctions can be very useful. In fact, the de Broglie plane wave $\Psi = \exp(ikx - i\omega t)$ for a free particle is a good example: since $|\Psi|^2 = 1$ the integral is in fact infinite. What this means is that $\exp(ikx - i\omega t)$ does not truly represent a single particle. To construct a square-integrable wavefunction we can use a superposition of plane waves. It is indeed a pleasant surprise that the superposition of infinitely many non-square integrable waves is square integrable!

2 The Wavefunction as a Probability Amplitude

Let's begin with a normalized wavefunction at initial time t_0

$$\int_{-\infty}^{\infty} \Psi^*(x, t_0) \Psi(x, t_0) dx = 1. \quad (2.1)$$

Since $\Psi(x, t_0)$ and the Schrödinger equation determine Ψ for all times, do we then have

$$\int_{-\infty}^{\infty} \Psi^*(x, t) \Psi(x, t) dx = 1 ? \quad (2.2)$$

Define the **probability density** $\rho(x, t)$

$$\rho(x, t) \equiv \Psi^*(x, t) \Psi(x, t) = |\Psi(x, t)|^2. \quad (2.3)$$

Define also $\mathcal{N}(t)$ as the integral of the probability density throughout space:

$$\mathcal{N}(t) \equiv \int \rho(x, t) dx. \quad (2.4)$$

The statement in (2.1) that the wavefunction begins well normalized is

$$\mathcal{N}(t_0) = 1, \quad (2.5)$$

and the condition that it remain normalized for all later times is $\mathcal{N}(t) = 1$. This would be guaranteed if we showed that for all times

$$\frac{d\mathcal{N}(t)}{dt} = 0. \quad (2.6)$$

We call this *conservation* of probability. Let's check if the Schrödinger equation ensures this condition will hold:

$$\begin{aligned} \frac{d\mathcal{N}(t)}{dt} &= \int_{-\infty}^{\infty} \frac{\partial \rho(x, t)}{\partial t} dx \\ &= \int_{-\infty}^{\infty} \left(\frac{\partial \Psi^*}{\partial t} \Psi(x, t) + \Psi^*(x, t) \frac{\partial \Psi(x, t)}{\partial t} \right) dx. \end{aligned} \quad (2.7)$$

From the Schrödinger equation, and its complex conjugate

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi \implies \frac{\partial \Psi}{\partial t} = -\frac{i}{\hbar} \hat{H}\Psi, \quad (2.8)$$

$$-i\hbar \frac{\partial \Psi^*}{\partial t} = (\hat{H}\Psi)^* \implies \frac{\partial \Psi^*}{\partial t} = \frac{i}{\hbar} (\hat{H}\Psi)^*. \quad (2.9)$$

In complex conjugating the Schrödinger equation we used that the complex conjugate of the time derivative of Ψ is simply the time derivative of the complex conjugate of Ψ . To conjugate

the right hand side we simply added the star to the whole of $\hat{H}\Psi$. We now use (2.8) and (2.9) in (2.7) to find

$$\begin{aligned}\frac{d\mathcal{N}(t)}{dt} &= \int_{-\infty}^{\infty} \left(\frac{i}{\hbar}(\hat{H}\Psi)^* \Psi - \frac{i}{\hbar}\Psi^*(\hat{H}\Psi) \right) dx \\ &= \frac{i}{\hbar} \left(\int_{-\infty}^{\infty} (\hat{H}\Psi)^* \Psi dx - \int_{-\infty}^{\infty} \Psi^*(\hat{H}\Psi) dx \right).\end{aligned}\quad (2.10)$$

To show that the time derivative of $\mathcal{N}(t)$ vanishes, it suffices to show that

$$\boxed{\int_{-\infty}^{\infty} (\hat{H}\Psi)^* \Psi = \int_{-\infty}^{\infty} \Psi^*(\hat{H}\Psi).}\quad (2.11)$$

Equation (2.11) is the condition on the Hamiltonian operator \hat{H} for conservation of probability. In fact, if \hat{H} is a Hermitian operator the condition will be satisfied. The operator \hat{H} is a **Hermitian** operator if it satisfies

$$\boxed{\text{Hermitian operator: } \int_{-\infty}^{\infty} (\hat{H}\Psi_1)^* \Psi_2 = \int_{-\infty}^{\infty} \Psi_1^*(\hat{H}\Psi_2).}\quad (2.12)$$

Here we have two wavefunctions that are arbitrary, but satisfy the conditions (1.5) and (1.6). As you can see, a Hermitian operator can be switched from acting on the first function to acting on the second function. When the two functions are the same, we recover condition (2.11).

It is worth closing this circle of ideas by defining the **Hermitian conjugate** T^\dagger of the linear operator T . This is done as follows:

$$\int_{-\infty}^{\infty} \Psi_1^*(T\Psi_2) = \int_{-\infty}^{\infty} (T^\dagger \Psi_1)^* \Psi_2.\quad (2.13)$$

The operator T^\dagger , which is also linear, is calculated by starting from the left-hand side and trying to recast the expression with no operator acting on the second function. An operator T is said to be Hermitian if it is equal to its Hermitian conjugate:

$$T \text{ is Hermitian if } T^\dagger = T.\quad (2.14)$$

Hermitian operators are very important in Quantum Mechanics. They have real eigenvalues and one can always find a basis of the state space in terms of orthonormal eigenstates. It turns out that observables in Quantum Mechanics are represented by Hermitian operators, and the possible measured values of those observables are given by their eigenvalues. Our quest to show that normalization is preserved under time evolution in Quantum Mechanics has come down to showing that the Hamiltonian operator is Hermitian.

3 The Probability Current

Let's take a closer look at the integrand of equation (2.10). Using the explicit expression for the Hamiltonian we have

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= \frac{i}{\hbar}((\hat{H}\Psi)^* \Psi - \Psi^*(\hat{H}\Psi)) \\ &= \frac{i}{\hbar} \left[-\frac{\hbar^2}{2m} \left(\frac{\partial^2 \Psi^*}{\partial x^2} \Psi - \Psi^* \frac{\partial^2 \Psi}{\partial x^2} \right) + V(x, t) \Psi^* \Psi - \Psi^* V(x, t) \Psi \right].\end{aligned}\quad (3.1)$$

The contributions from the potential cancel and we then get

$$\frac{i}{\hbar}((\hat{H}\Psi)^* \Psi - \Psi^*(\hat{H}\Psi)) = \frac{\hbar}{2im} \left(\frac{\partial^2 \Psi^*}{\partial x^2} \Psi - \Psi^* \frac{\partial^2 \Psi}{\partial x^2} \right). \quad (3.2)$$

The only chance to get to show that the integral of the right-hand side is zero is to show that it is a total derivative. Indeed, it is!

$$\begin{aligned}\frac{i}{\hbar}((\hat{H}\Psi)^* \Psi - \Psi^*(\hat{H}\Psi)) &= \frac{\partial}{\partial x} \left[\frac{\hbar}{2im} \left(\frac{\partial \Psi^*}{\partial x} \Psi - \Psi^* \frac{\partial \Psi}{\partial x} \right) \right] \\ &= -\frac{\partial}{\partial x} \left[\frac{\hbar}{2im} \left(\Psi^* \frac{\partial \Psi}{\partial x} - \frac{\partial \Psi^*}{\partial x} \Psi \right) \right] \\ &= -\frac{\partial}{\partial x} \left[\frac{\hbar}{2im} 2i \operatorname{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right) \right] \\ &= -\frac{\partial}{\partial x} \left[\frac{\hbar}{m} \operatorname{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right) \right],\end{aligned}\quad (3.3)$$

where we used that $z - z^* = 2i \operatorname{Im}(z)$. Recall that the left-hand side we have evaluated is actually $\frac{\partial \rho}{\partial t}$ and therefore the result obtained so far is

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} \left[\frac{\hbar}{m} \operatorname{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right) \right] = 0. \quad (3.4)$$

This equation encodes charge conservation and is of the type

$$\frac{\partial \rho}{\partial t} + \frac{\partial J}{\partial x} = 0, \quad (3.5)$$

where $J(x, t)$ is the current associated with the charge density ρ . We have therefore identified a probability current

$$J(x, t) \equiv \frac{\hbar}{m} \operatorname{Im} \left(\Psi^* \frac{\partial \Psi}{\partial x} \right).$$

(3.6)

There is just one component for this current since the particle moves in one dimension. The units of J are one over time, or probability per unit time, as we now verify.

For one spatial dimension, $[\Psi] = L^{-1/2}$, which is easily seen from the requirement that $\int dx |\Psi|^2$ is unit free. (When working with d spatial dimensions the wavefunction will have units of $L^{-d/2}$). We then have

$$\left[\Psi^* \frac{\partial \Psi}{\partial x} \right] = \frac{1}{L^2}, \quad [\hbar] = \frac{ML^2}{T}, \quad \left[\frac{\hbar}{m} \right] = \frac{L^2}{T}, \quad (3.7)$$

$$\implies [J] = \frac{1}{T} = \text{probability per unit time} \quad (3.8)$$

We can now show that the time derivative of \mathcal{N} is zero. Indeed, using (3.5) we have

$$\frac{d\mathcal{N}}{dt} = \int_{-\infty}^{\infty} dx \frac{\partial \rho}{\partial t} = - \int_{-\infty}^{\infty} \frac{\partial J}{\partial x} dx = -(J(\infty, t) - J(-\infty, t)). \quad (3.9)$$

The derivative vanishes if the probability current vanishes at infinity. Recalling that

$$J = \frac{\hbar}{2im} \left(\Psi^* \frac{\partial \Psi}{\partial x} - \Psi \frac{\partial \Psi^*}{\partial x} \right), \quad (3.10)$$

we see that the current indeed vanishes because we restrict ourselves to wavefunctions for which $\lim_{x \rightarrow \pm\infty} \Psi = 0$ and $\lim_{x \rightarrow \pm\infty} \frac{\partial \Psi}{\partial x}$ remains bounded. We therefore have

$$\frac{d\mathcal{N}}{dt} = 0, \quad (3.11)$$

as we wanted to show.

To illustrate how probability conservation works more generally in one dimension, focus on a segment $x \in [a, b]$. Then the probability P_{ab} to find the particle in the segment $[a, b]$, is given by

$$P_{ab} = \int_a^b \rho(x, t) dx. \quad (3.12)$$

If we now take the time derivative of this and, as before, use current conservation we get

$$\frac{dP_{ab}}{dt} = - \int_a^b \frac{\partial J(x, t)}{\partial x} dt = -J(b, t) + J(a, t). \quad (3.13)$$

This is the expected result. If the amount of probability in the region $[a, b]$ changes in time, it must be due to probability current flowing in or out at the edges of the interval. Assuming the currents at $x = b$ and at $x = a$ are positive, we note that probability is flowing out at $x = b$ and is coming in at $x = a$. The signs in the above right-hand side correctly reflect the effect of these flows on the rate of change of the total probability inside the segment.

4 Probability current in 3D and current conservation

The determination of the probability current \mathbf{J} for a particle moving in three dimensions follows the route taken before, but we use the 3D version of the Schrödinger equation. After some work (homework) the probability density and the current are determined to be

$$\rho(\mathbf{x}, t) = |\Psi(\mathbf{x}, t)|^2, \quad \mathbf{J}(x, t) = \frac{\hbar}{m} \operatorname{Im} (\Psi^* \nabla \Psi), \quad (4.1)$$

and satisfy the conservation equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0. \quad (4.2)$$

In three spatial dimensions, $[\Psi] = L^{-\frac{3}{2}}$ and the units of \mathbf{J} are quickly determined

$$[\Psi^* \nabla \Psi] = \frac{1}{L^4}, \quad \left[\frac{\hbar}{m} \right] = \frac{L^2}{T} \quad (4.3)$$

$$\implies [\mathbf{J}] = \frac{1}{TL^2} = \text{probability per unit time per unit area} \quad (4.4)$$

The conservation equation (4.2) is particularly clear in integral language. Consider a fixed region V of space and the probability $Q_V(t)$ to find the particle inside the region:

$$Q_V(t) = \int_V \rho(\mathbf{x}, t) d^3 \mathbf{x}. \quad (4.5)$$

The time derivative of the probability is then calculated using the conservation equation

$$\frac{dQ_V}{dt} = \int_V \frac{\partial \rho}{\partial t} d^3 \mathbf{x} = - \int_V \nabla \cdot \mathbf{J} d^3 \mathbf{x}. \quad (4.6)$$

Finally, using Gauss' law we find

$$\frac{dQ_V}{dt} = - \int_S \mathbf{J} \cdot d\mathbf{a}, \quad (4.7)$$

where S is the boundary of the volume V . The interpretation here is clear. The probability that the particle is inside V may change in time if there is flux of the probability current across the boundary of the region. When the volume extends throughout space, the boundary is at infinity, and the conditions on the wavefunction (which we have not discussed in the 3D case) imply that the flux across the boundary at infinity vanishes.

Our probability density, probability current, and current conservation are in perfect analogy to electromagnetic charge density, current density, and current conservation. In electromagnetism charges flow, in quantum mechanics probability flows. The terms of the correspondence are summarized by the table.

	Electromagnetism	Quantum Mechanics
ρ	charge density	probability density
Q_V	charge in a volume V	probability to find particle in V
\mathbf{J}	current density	probability current density

Sarah Geller transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare
<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

Lectures 21 and 22: Hydrogen Atom

B. Zwiebach
May 4, 2016

Contents

1 The Hydrogen Atom	1
2 Hydrogen atom spectrum	4

1 The Hydrogen Atom

Our goal here is to show that the two-body quantum mechanical problem of the hydrogen atom can be recast as one in which we have center-of-mass degrees of freedom that behave like a free particle and relative-motion degrees of freedom for which we have dynamics controlled by a central potential.

The hydrogen atom consists of a proton and an electron moving in three dimensions. We label the position and momentum operators of the proton as $\hat{\mathbf{x}}_p, \hat{\mathbf{p}}_p$, and those of the electron as $\hat{\mathbf{x}}_e, \hat{\mathbf{p}}_e$. These are canonical variables, meaning they satisfy the canonical commutation relations:

$$[(\hat{\mathbf{x}}_p)_i, (\hat{\mathbf{p}}_p)_j] = i\hbar\delta_{ij}, \quad [(\hat{\mathbf{x}}_e)_i, (\hat{\mathbf{p}}_e)_j] = i\hbar\delta_{ij}. \quad (1.1)$$

Here the subscripts $i, j = 1, 2, 3$ denote the various components of the vector operators. Furthermore, the proton variables **commute** with the electron variables. We have two pairs of *independent* canonical variables.

The wavefunction for the system is a function of the positions of both particles:

$$\Psi(\mathbf{x}_p, \mathbf{x}_e), \quad (1.2)$$

and the quantity

$$|\Psi(\mathbf{x}_p, \mathbf{x}_e)|^2 d^3\mathbf{x}_p d^3\mathbf{x}_e, \quad (1.3)$$

is the probability to find the proton within a window $d^3\mathbf{x}_p$ of \mathbf{x}_p and the electron within a window $d^3\mathbf{x}_e$ of \mathbf{x}_e . The Hamiltonian of the system is given by

$$\hat{H} = \frac{\hat{\mathbf{p}}_p^2}{2m_p} + \frac{\hat{\mathbf{p}}_e^2}{2m_e} + V(|\mathbf{x}_e - \mathbf{x}_p|). \quad (1.4)$$

Note that the kinetic energy is simply the sum of the kinetic energy of the proton and kinetic energy of the electron. The potential only depends on the magnitude of the separation between the two particles, not on their individual positions.

In order to simplify the problem, we will introduce two new pairs of independent canonical variables. The first pair is associated with the center-of-mass (CM) motion. We introduce the total momentum operator $\hat{\mathbf{P}}$ and the CM position operator $\hat{\mathbf{X}}$, given by

$$\hat{\mathbf{P}} = \hat{\mathbf{p}}_p + \hat{\mathbf{p}}_e, \quad \hat{\mathbf{X}} = \frac{m_e \hat{\mathbf{x}}_e + m_p \hat{\mathbf{x}}_p}{m_e + m_p}. \quad (1.5)$$

The operator $\hat{\mathbf{X}}$ is given by the typical expression for the center-of-mass of the system, but with the positions replaced by position operators. Using the commutation relations (1.1), we can show that $\hat{\mathbf{X}}$ and $\hat{\mathbf{P}}$ are canonical conjugates:

$$\begin{aligned} [(\hat{\mathbf{X}})_i, (\hat{\mathbf{P}})_j] &= \left[\frac{m_e(\hat{\mathbf{x}}_e)_i + m_p(\hat{\mathbf{x}}_p)_i}{m_e + m_p}, (\hat{\mathbf{p}}_p)_j + (\hat{\mathbf{p}}_e)_j \right] \\ &= \frac{m_e}{m_e + m_p} [(\hat{\mathbf{x}}_e)_i, (\hat{\mathbf{p}}_e)_j] + \frac{m_p}{m_e + m_p} [(\hat{\mathbf{x}}_p)_i, (\hat{\mathbf{p}}_p)_j] \\ &= \frac{m_e}{m_e + m_p} i\hbar\delta_{ij} + \frac{m_p}{m_e + m_p} i\hbar\delta_{ij}, \end{aligned} \quad (1.6)$$

resulting in the expected

$$[(\hat{\mathbf{X}})_i, (\hat{\mathbf{P}})_j] = i\hbar\delta_{ij}. \quad (1.7)$$

For the second pair of canonical variables we will define relative position and momentum operators. The relative position operator is the natural variable implied by the form of the potential:

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_e - \hat{\mathbf{x}}_p. \quad (1.8)$$

Since the second pair of canonical variables must commute with the first pair, we must check that \mathbf{x} defined above, commutes with \mathbf{X} and with \mathbf{P} . The commutation with \mathbf{X} is automatic and the commutation with \mathbf{P} works thanks to the minus sign in the above definition. We must now construct a relative momentum operator $\hat{\mathbf{p}}$ that is canonically conjugate to \mathbf{x} . It must be built from the momentum operators of the two particles, so we write

$$\hat{\mathbf{p}} = \alpha \hat{\mathbf{p}}_e - \beta \hat{\mathbf{p}}_p, \quad (1.9)$$

with α and β coefficients to be determined. To be canonically conjugate, the relative operators must satisfy

$$[(\hat{\mathbf{x}})_i, (\hat{\mathbf{p}})_j] = i\hbar\delta_{ij} \rightarrow \alpha + \beta = 1, \quad (1.10)$$

using the above definitions of $\hat{\mathbf{x}}$ and $\hat{\mathbf{p}}$ and the proton and electron commutators. Finally, the relative momentum must commute with the CM coordinate

$$[(\hat{\mathbf{X}})_i, (\hat{\mathbf{p}})_j] = 0 \rightarrow m_e\alpha - m_p\beta = 0. \quad (1.11)$$

The two equations for α and β can be solved to find

$$\alpha = \frac{m_p}{m_e + m_p}, \quad \beta = \frac{m_e}{m_e + m_p}. \quad (1.12)$$

We define the total mass M and the reduced mass μ as follows

$$M = m_e + m_p, \quad \mu = \frac{m_e m_p}{m_e + m_p}. \quad (1.13)$$

The reduced mass of a pair of particles with very different masses is approximately equal to the mass of the lower-mass particle. Using these definitions

$$\alpha = \frac{\mu}{m_e}, \quad \beta = \frac{\mu}{m_p}. \quad (1.14)$$

Thus, collecting the relative variables we have

$$\hat{\mathbf{p}} = \mu \left(\frac{\hat{\mathbf{p}}_e}{m_e} - \frac{\hat{\mathbf{p}}_p}{m_p} \right) = \frac{m_p}{M} \hat{\mathbf{p}}_e - \frac{m_e}{M} \hat{\mathbf{p}}_p, \quad \hat{\mathbf{x}} = \hat{\mathbf{x}}_e - \hat{\mathbf{x}}_p. \quad (1.15)$$

Note that the relative momentum \mathbf{p} can be written in terms of velocities as follows: $\mathbf{p} = \mu(\mathbf{v}_e - \mathbf{v}_p)$. The relative momentum vanishes if the motion is only CM motion, in which case the velocities of the two parties are the same.

We can now rewrite the Hamiltonian in terms of the new variables. Solving for the original momentum operators in terms of $\hat{\mathbf{P}}$ and $\hat{\mathbf{p}}$, we find

$$\hat{\mathbf{p}}_p = \frac{m_p}{M} \hat{\mathbf{P}} - \hat{\mathbf{p}}, \quad \hat{\mathbf{p}}_e = \frac{m_e}{M} \hat{\mathbf{P}} + \hat{\mathbf{p}}. \quad (1.16)$$

We can then rewrite the kinetic terms of the Hamiltonian in the form

$$\begin{aligned} \frac{\hat{\mathbf{p}}_p^2}{2m_p} + \frac{\hat{\mathbf{p}}_e^2}{2m_e} &= \frac{1}{2m_p} \left(\frac{m_p^2}{M^2} \hat{\mathbf{P}}^2 - \frac{2m_p}{M} \hat{\mathbf{P}} \cdot \hat{\mathbf{p}} + \hat{\mathbf{p}}^2 \right) \\ &\quad + \frac{1}{2m_e} \left(\frac{m_e^2}{M^2} \hat{\mathbf{P}}^2 + \frac{2m_e}{M} \hat{\mathbf{P}} \cdot \hat{\mathbf{p}} + \hat{\mathbf{p}}^2 \right) \\ &= \frac{\hat{\mathbf{P}}^2}{2M} + \frac{\hat{\mathbf{p}}^2}{2\mu}. \end{aligned} \quad (1.17)$$

Happily the term coupling the two momenta vanishes. Thus the center of mass degrees of freedom and the relative degrees of freedom give independent contributions to the kinetic energy. The Hamiltonian can then be written as

$$\hat{H} = \frac{\hat{\mathbf{P}}^2}{2M} + \frac{\hat{\mathbf{p}}^2}{2\mu} + V(\hat{\mathbf{x}}). \quad (1.18)$$

In position space, the total and relative momentum operators can be expressed as gradients

$$\hat{\mathbf{P}} \rightarrow \frac{\hbar}{i} \nabla_{\mathbf{X}}, \quad \hat{\mathbf{p}} \rightarrow \frac{\hbar}{i} \nabla_{\mathbf{x}}. \quad (1.19)$$

Each ∇ has a subscript indicating the type of coordinate we use to take the derivatives. Just like we had a wavefunction $\Psi(\mathbf{x}_e, \mathbf{x}_p)$ the new canonical variables require that we now think of the wavefunction as a function $\Psi(\mathbf{X}, \mathbf{x})$ of the new coordinates.

We solve the time-independent Schrödinger equation by using separation of variables

$$\Psi(\mathbf{X}, \mathbf{x}) = \Psi_{\text{CM}}(\mathbf{X}) \Psi_{\text{rel}}(\mathbf{x}). \quad (1.20)$$

Plugging these into the time-independent Schrödinger equation $\hat{H}\Psi = E\Psi$, we reach

$$\left[\frac{\hat{\mathbf{P}}^2}{2M} \Psi_{\text{CM}}(\mathbf{X}) \right] \Psi_{\text{rel}}(\mathbf{x}) + \left[\frac{\hat{\mathbf{p}}^2}{2\mu} \Psi_{\text{rel}}(\mathbf{x}) + V(|\hat{\mathbf{x}}|) \Psi_{\text{rel}}(\mathbf{x}) \right] \Psi_{\text{CM}}(\mathbf{X}) = E \Psi_{\text{CM}}(\mathbf{X}) \Psi_{\text{rel}}(\mathbf{x}). \quad (1.21)$$

Dividing by the total wavefunction $\Psi_{\text{CM}}(\mathbf{X}) \Psi_{\text{rel}}(\mathbf{x})$, this becomes

$$\frac{1}{\Psi_{\text{CM}}(\mathbf{X})} \left[\frac{\hat{\mathbf{P}}^2}{2M} \Psi_{\text{CM}}(\mathbf{X}) \right] + \frac{1}{\Psi_{\text{rel}}(\mathbf{x})} \left[\frac{\hat{\mathbf{p}}^2}{2\mu} + V(|\hat{\mathbf{x}}|) \right] \Psi_{\text{rel}}(\mathbf{x}) = E. \quad (1.22)$$

The first term on the left-hand side is a function of \mathbf{X} only and the second term on the left-hand side is a function of \mathbf{x} only. Their sum is equal to the constant E and since \mathbf{x} and \mathbf{X} are independent variables, each term must individually be constant. We thus set the first term equal to the constant E_{CM} and the second term equal to the constant E_{rel} , resulting in the following equations:

$$\frac{\hat{\mathbf{P}}^2}{2M} \Psi_{\text{CM}}(\mathbf{X}) = E_{\text{CM}} \Psi_{\text{CM}}(\mathbf{X}), \quad (1.23)$$

$$\left[\frac{\hat{\mathbf{p}}^2}{2\mu} + V(|\mathbf{x}|) \right] \Psi_{\text{rel}}(\mathbf{x}) = E_{\text{rel}} \Psi_{\text{rel}}(\mathbf{x}), \quad (1.24)$$

$$E = E_{\text{CM}} + E_{\text{rel}}. \quad (1.25)$$

We get two Schrödinger equations. The first equation tells us that the center of mass moves as a free particle of mass M . Thus, the CM energy is not quantized and we get plane wave solutions. The second equation is for the relative motion, and as we wanted to show, it is described as motion in a central potential. The third equation tells us that the total energy is the sum of the center-of-mass energy and the energy from the relative motion.

2 Hydrogen atom spectrum

We now have the tools to study the hydrogen atom, which has a central potential given by

$$V(r) = -\frac{Ze^2}{r}, \quad (2.1)$$

where Z is the number of protons in the nucleus. For hydrogen we have $Z = 1$. But it is worth considering $Z > 1$ in which case we are describing the motion of an electron around the nucleus of some other atom. We will also define following physical constants.

- The fine structure constant α : $\alpha = \frac{e^2}{\hbar c} \simeq \frac{1}{137}$.
- The *Bohr radius* a_0 . This is the characteristic length scale in the problem. It can be calculated by equating kinetic and potential energies expressed in terms of a_0 and ignoring all numerical constants:

$$\frac{\hbar^2}{m_e a_0^2} = \frac{e^2}{a_0}. \quad (2.2)$$

Here the mass should be the reduced mass, which in this case can be taken rather accurately to be the mass of the electron. We then have explicitly,

$$\begin{aligned} a_0 &= \frac{\hbar^2}{me^2} = \frac{\hbar^2 c^2}{e^2 mc^2} = \frac{\hbar c}{\left(\frac{e^2}{\hbar c}\right) mc^2} = \frac{\hbar c}{\alpha m c^2} \\ &= \frac{197 \text{ MeV fm}}{0.51 \times 10^6 \text{ eV} \left(\frac{1}{137}\right)} = \frac{1970 \text{ eV Angstrom}}{0.51 \times 10^6 \text{ eV}} \times 137 \\ &= 0.529 \text{ Angstroms} \simeq 53 \text{ pm}. \end{aligned} \quad (2.3)$$

For the energy scale estimate we have

$$\frac{e^2}{a_0} = e^2 \left(\frac{me^2}{\hbar^2} \right) = \left(\frac{e^4}{\hbar^2 c^2} \right) mc^2 = \alpha^2 mc^2 = \frac{1}{(137)^2} \times (511\,000 \text{ eV}) \simeq 27.2 \text{ eV} \quad (2.4)$$

There are other characteristic lengths that are interesting:

$$\begin{aligned}\alpha a_0 &= \text{Compton wavelength of electron} = \lambda_e \simeq 390 \text{ fm}, \\ \alpha^2 a_0 &= \text{classical electron radius} \simeq 2.8 \text{ fm}.\end{aligned}\quad (2.5)$$

Let us do some work now! The radial Schrödinger equation for bound states $E < 0$ reads

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dr^2} + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} - \frac{Ze^2}{r} \right) u = Eu. \quad (2.6)$$

We could label the wavefunction u as $u_{E\ell}$ as the solutions will certainly depend on ℓ and the energy E . As usual, we like to work with a unit free coordinate. This could be achieved by writing $r = a_0 x$, with x unit free and a_0 carrying the length units of r . It will be more convenient to use a slight variation to eliminate Z from the equation and some factors of two. We will take the new unit-free coordinate x to be defined from

$$r \equiv \frac{a_0}{2Z} x. \quad (2.7)$$

The Schrödinger equation then becomes

$$\begin{aligned}&\left(-\frac{\hbar^2}{2m} \frac{4Z^2}{a_0^2} \frac{d^2}{dx^2} + \frac{4Z^2}{a_0^2} \frac{\hbar^2}{2m} \frac{l(l+1)}{x^2} - \frac{2Z^2 e^2}{a_0} \frac{1}{x} \right) u = Eu \\ &\rightarrow \left(\frac{2\hbar^2 Z^2}{ma_0^2} \left(-\frac{d^2}{dx^2} + \frac{l(l+1)}{x^2} \right) - \frac{2Z^2 e^2}{a_0} \frac{1}{x} \right) u = Eu.\end{aligned}\quad (2.8)$$

Note that

$$\frac{2\hbar^2 Z^2}{ma_0^2} = \frac{2\hbar^2 Z}{ma_0} \frac{me^2}{\hbar^2} = \frac{2Ze^2}{a_0}, \quad (2.9)$$

which reduces our differential equation to

$$\left(-\frac{d^2}{dx^2} + \frac{\ell(\ell+1)}{x^2} - \frac{1}{x} \right) u = \frac{E}{\left(\frac{2Ze^2}{a_0} \right)} u. \quad (2.10)$$

We now define the unit-free parameter κ that encodes the energy:

$$\kappa^2 = -\frac{E}{\left(\frac{2Ze^2}{a_0} \right)} > 0. \quad (2.11)$$

κ is a unit-free version of the bound state energy. The differential equation is then

$$\left(-\frac{d^2}{dx^2} + \frac{\ell(\ell+1)}{x^2} - \frac{1}{x} \right) u = -\kappa^2 u. \quad (2.12)$$

We can further simplify this equation prior to solving it by examining the limiting cases. In the limit $x \rightarrow \infty$, the dominant terms are the second derivative and the term on the right-hand side, giving

$$\frac{d^2 u}{dx^2} = \kappa^2 u \implies u \sim e^{\pm \kappa x}. \quad (2.13)$$

Since κ is unit free, we can make the above exponent equal to a new unit-free coordinate ρ :

$$\boxed{\rho \equiv \kappa x = \frac{2\kappa Z}{a_0} r.} \quad (2.14)$$

This time we get

$$\boxed{\left(-\frac{d^2}{d\rho^2} + \frac{\ell(\ell+1)}{\rho^2} - \frac{1}{\kappa\rho} \right) u = -u.} \quad (2.15)$$

Note that we did not get κ to disappear from the equation. This is good news: the equation should fix the possible values of κ (or possible energies). The equation above is not quite ready for a series solution: we would find a three term recursion relation, which is rather complicated. To make progress we discuss the behavior for small and large ρ .

For $\rho \rightarrow \infty$ we now get $u \sim u^{\pm\rho}$ and of course we hope for $u = e^{-\rho}$ for normalizability. As we discussed before, for $\rho \rightarrow 0$ the radial solution must be of the form $u \sim \rho^{(l+1)}$. This information about the behavior for small and for large ρ suggests a good ansatz for $u(\rho)$

$$\boxed{u(\rho) = \rho^{\ell+1} W(\rho) e^{-\rho}.} \quad (2.16)$$

where $W(\rho)$ is a yet to be determined function that we hope satisfies a simpler differential equation. To derive this differential equation for $W(\rho)$, we plug our ansatz into Eq. (2.15). As a little help on the calculation, we give an intermediate result:

$$-u'' + \frac{\ell(\ell+1)}{\rho^2} u + u = \left(-W'' - \frac{2(\ell+1)}{\rho} W' + \frac{2(\ell+1)}{\rho} W + 2W' \right) \rho^{\ell+1} e^{-\rho}. \quad (2.17)$$

With a little more work we finally get the differential equation for W :

$$\boxed{\rho \frac{d^2W}{d\rho^2} + 2(\ell+1-\rho) \frac{dW}{d\rho} + \left[\frac{1}{\kappa} - 2(\ell+1) \right] W = 0.} \quad (2.18)$$

This looks a bit more complicated than the differential equation we started with but it leads to a very nice one-step recursion relation. As usual we write W as a series expansion

$$W = \sum_{k=0}^{\infty} a_k \rho^k, \quad (2.19)$$

and plugging back into (2.18), we group terms of order ρ^k to derive a recursion relation

$$\begin{aligned} a_{k+1} k(k+1) + 2(\ell+1)(k+1)a_{k+1} - 2ka_k + \left[\frac{1}{\kappa} - 2(\ell+1) \right] a_k &= 0, \\ \rightarrow a_{k+1}(k(k+1) + 2(\ell+1)(k+1)) &= a_k \left(2(\ell+k+1) - \frac{1}{\kappa} \right), \end{aligned} \quad (2.20)$$

which gives

$$\frac{a_{k+1}}{a_k} = \frac{2(k+\ell+1) - \frac{1}{\kappa}}{(k+1)(k+2\ell+2)}. \quad (2.21)$$

Detailed examination shows that for normalizable wave functions, the series must terminate. To see this we examine the large k behavior of the above ratio:

$$\frac{a_{k+1}}{a_k} \simeq \frac{2k}{k^2} = \frac{2}{k}. \quad (2.22)$$

Note that $\frac{2}{k+1} < \frac{2}{k}$; thus if the ratio $\frac{2}{k+1}$ leads to a divergence so will the ratio $\frac{2}{k}$. Taking

$$\frac{a_{k+1}}{a_k} = \frac{2}{k+1} \rightarrow a_{k+1} = \frac{2}{k+1} a_k, \quad (2.23)$$

and this is solved by

$$a_k = \frac{2^k}{k!} a_0. \quad (2.24)$$

Therefore, the sum

$$W = \sum_{k=0}^{\infty} a_k \rho^k \simeq a_0 \sum_{k=0}^{\infty} \frac{2^k \rho^k}{k!} = a_0 e^{2\rho}. \quad (2.25)$$

This is precisely sufficient to make the ansatz in (2.16) un-normalizable.

In order to get a normalizable solution the series for W must terminate. Suppose W is a polynomial of degree N so the coefficients satisfy

$$a_N \neq 0 \quad \text{and} \quad a_{N+1} = 0. \quad (2.26)$$

From Eq.(2.21) this implies

$$\frac{1}{\kappa} = 2(N + \ell + 1). \quad (2.27)$$

Quantization has happened! The energy-encoding parameter κ is now related to integers! Note that ℓ can take values $\ell = 0, 1, 2, \dots$ as it befits it being an angular momentum quantum number. Moreover N can take values $N = 0, 1, 2, \dots$, since a polynomial of degree zero exists, being equal to a constant. Define the **principal** quantum number n as follows:

$$n \equiv N + \ell + 1 = \frac{1}{2\kappa}, \quad \text{with } \ell = 0, 1, 2, \dots, \quad N = 0, 1, 2, \dots, \quad \text{and } n = 1, 2, 3, \dots. \quad (2.28)$$

Importantly, note that for a fixed n we must have

$$0 \leq \ell \leq n - 1, \quad \text{and} \quad 0 \leq N \leq n - 1. \quad (2.29)$$

If n and ℓ are known, N is determined from $N + \ell + 1 = n$. So the independent quantum numbers so far are n and ℓ . Interestingly, the energies depend only on n , since κ depends only on n . Using Eq. (2.11), the energy dependence on principle quantum number is given by

$$E = -\frac{2Z^2 e^2}{a_0} \kappa^2, \quad (2.30)$$

and using $\kappa = \frac{1}{2n}$ we get

$$E = -\frac{Z^2 e^2}{2a_0} \frac{1}{n^2}. \quad (2.31)$$

These are the energy levels of the hydrogen atom! Since at any fixed value of $n > 1$ there are various possible ℓ values, the spectrum is highly degenerate. Even more, each value of ℓ amounts to $2\ell + 1$ states, given the possible values of m . One way to visualize the spectrum is shown in Figure 1. All integer points in the (N, ℓ) positive quadrant represent states. The states with common value of n lie on the dashed lines.

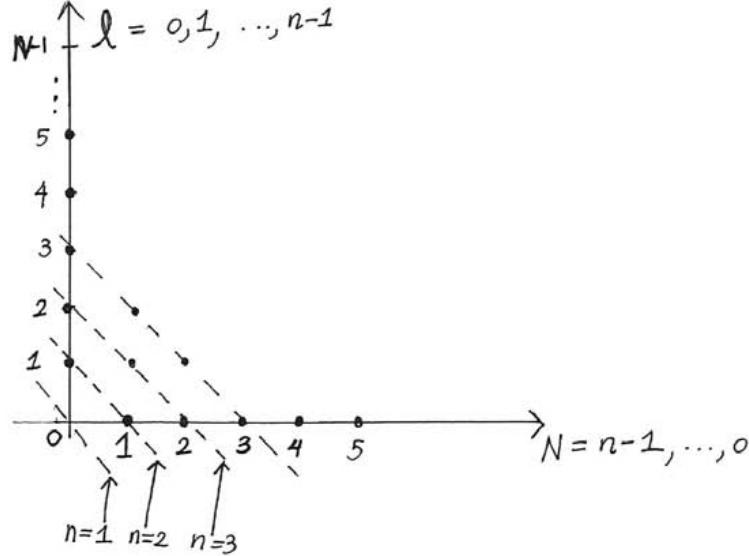


Figure 1: All points with integer $N \geq 0$ and integer $\ell \geq 0$ represent hydrogen atom states. The figure helps us to count the number of possible states for given value of n . Each dot along the diagonal line for a given n represents a possible state.

Figure 1 helps us to count the number of bound states for a given value of n . Recall that for each n , ℓ can take values from $0, \dots, n-1$ and for each value of ℓ , m takes values from $-\ell$ to ℓ . The following table counts the states for the first few values of the principal quantum number n . A given state is specified by its values for (n, ℓ, m) , all of which are referred to as the *quantum numbers* for hydrogen states. Each number has a very important physical meaning: n tells us about the energy eigenvalue, $\hbar^2 \ell(\ell + 1)$ is the eigenvalue of the square of angular momentum and $\hbar m$ is the eigenvalue of the z component of angular momentum.

Number of States

n value	ℓ values	m values	total states
$n=1$	$\ell=0$	$m=0$	1 state
$n=2$	$\ell=0$	$m=0$	1
	$\ell=1$	$m=-1, 0, 1$	+ 3 = 4 states
$n=3$	$\ell=0$	$m=0$	1
	$\ell=1$	$m=-1, 0, 1$	+ 3
	$\ell=2$	$m=-2, \dots, 2$	+ 5 = 9 states

The total number of states for arbitrary principal quantum number n can now be calculated:

$$\# \text{ of states for } n = \sum_{\ell=0}^{n-1} (2\ell + 1) = \frac{2(n-1)n}{2} + n = n^2 - n + n = n^2. \quad (2.32)$$

This is in agreement with the partial results in the table. A more familiar representation of the states of hydrogen is given in Figure 2. The different columns indicate the different values of ℓ . We have also indicated in the figure the values of N , the degree of the polynomial entering the radial solution. Note that for a given ℓ , that is, for a fixed radial equation, the value of N increases as we go up the column. The number N corresponds to the number of nodes in the solution.

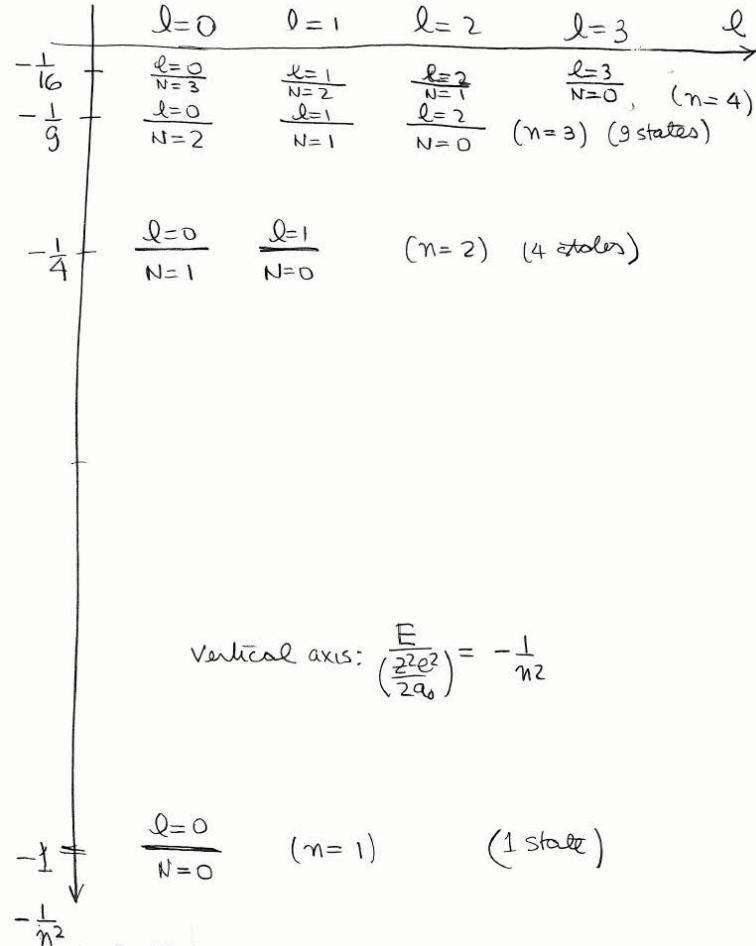


Figure 2: Plot of energy levels $E \sim -1/n^2$ indicating also the angular quantum number ℓ and the degree N of the polynomial. The spectrum is highly degenerate.

Recall that we defined $\rho = \frac{2\kappa Zr}{a_0}$. Together with $\kappa = \frac{1}{2n}$ this gives

$$\boxed{\rho = \frac{Zr}{na_0}.} \quad (2.33)$$

The eigenstates are labelled by quantum numbers (n, ℓ, m) and the wavefunctions are

$$\begin{aligned}\psi_{n\ell m} &= \mathcal{N} \frac{u_{n\ell}(r)}{r} Y_{\ell m}(\theta, \phi) = \mathcal{N} \frac{\rho^{\ell+1}}{\rho} W_{n\ell}(\rho) e^{-\rho} Y_{\ell m}(\theta, \phi) \\ &= \mathcal{N} \rho^\ell \underbrace{W_{n\ell}(\rho)}_{\substack{\text{polynomial of degree} \\ N=n-(\ell+1)}} e^{-\rho} Y_{\ell m}(\theta, \phi),\end{aligned}\tag{2.34}$$

where \mathcal{N} is a normalization constant. Therefore, using the expression for ρ and absorbing constants into \mathcal{N} we have

$$\psi_{n\ell m}(r, \theta, \phi) = \mathcal{N} \left(\frac{r}{a_0} \right)^\ell \left(\begin{array}{l} \text{polynomial in } \frac{r}{a_0} \\ \text{of degree } N=n-(\ell+1) \end{array} \right) e^{-\frac{Zr}{na_0}} Y_{\ell m}(\theta, \phi).\tag{2.35}$$

For the ground state of hydrogen ($Z = 1$), we have $(n, \ell, m) = (1, 0, 0)$. Having zero angular momentum the associated wavefunction has no angular dependence. The normalized wavefunction is

$$\psi_{100}(r, \theta, \phi) = \frac{1}{\sqrt{\pi a_0^3}} e^{-r/a_0}.\tag{2.36}$$

For normalized hydrogen wavefunctions at $n = 2$ and $n = 3$ see

<http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/hydwf.html>

Sarah Geller and Andrew Turner transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

MIT OpenCourseWare
<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

ANGULAR MOMENTUM

B. Zwiebach
December 16, 2013

Contents

1	Orbital angular momentum and central potentials	1
1.1	Quantum mechanical vector identities	2
1.2	Properties of angular momentum	6
1.3	The central potential Hamiltonian	9
2	Algebraic theory of angular momentum	11
3	Comments on spherical harmonics	18
4	The radial equation	20
5	The free particle and the infinite spherical well	24
5.1	Free particle	24
5.2	The infinite spherical well	25
6	The three-dimensional isotropic oscillator	28
7	Hydrogen atom and Runge-Lenz vector	33

1 Orbital angular momentum and central potentials

Classically the angular momentum vector \vec{L} is defined as the cross-product of the position vector \vec{r} and the momentum vector \vec{p} :

$$\vec{L} = \vec{r} \times \vec{p}. \quad (1.1)$$

In cartesian components, this equation reads

$$\begin{aligned} L_x &= yp_z - zp_y, \\ L_y &= zp_x - xp_z, \\ L_z &= xp_y - yp_x. \end{aligned} \quad (1.2)$$

In quantum mechanics the classical vectors \vec{r} , \vec{p} and \vec{L} become operators. More precisely, they give us triplets of operators:

$$\begin{aligned}\vec{r} &\rightarrow (\hat{x}, \hat{y}, \hat{z}), \\ \vec{p} &\rightarrow (\hat{p}_x, \hat{p}_y, \hat{p}_z), \\ \vec{L} &\rightarrow (\hat{L}_x, \hat{L}_y, \hat{L}_z).\end{aligned}\tag{1.3}$$

When we want more uniform notation, instead of x , y , and z labels we use 1, 2 and 3 labels:

$$\begin{aligned}(\hat{x}_1, \hat{x}_2, \hat{x}_3) &\equiv (\hat{x}, \hat{y}, \hat{z}), \\ (\hat{p}_1, \hat{p}_2, \hat{p}_3) &\equiv (\hat{p}_x, \hat{p}_y, \hat{p}_z), \\ (\hat{L}_1, \hat{L}_2, \hat{L}_3) &\equiv (\hat{L}_x, \hat{L}_y, \hat{L}_z).\end{aligned}\tag{1.4}$$

The basic canonical commutation relations then are easily summarized as

$$[\hat{x}_i, \hat{p}_j] = i\hbar \delta_{ij}, \quad [\hat{x}_i, \hat{x}_j] = 0, \quad [\hat{p}_i, \hat{p}_j] = 0.\tag{1.5}$$

Thus, for example, \hat{x} commutes with $\hat{y}, \hat{z}, \hat{p}_y$ and \hat{p}_z , but fails to commute with \hat{p}_x . In view of (1.2) and (1.3) it is natural to define the angular momentum *operators* by

$$\begin{aligned}\hat{L}_x &\equiv \hat{y}\hat{p}_z - \hat{z}\hat{p}_y, \\ \hat{L}_y &\equiv \hat{z}\hat{p}_x - \hat{x}\hat{p}_z, \\ \hat{L}_z &\equiv \hat{x}\hat{p}_y - \hat{y}\hat{p}_x.\end{aligned}\tag{1.6}$$

Note that these equations are free of ordering ambiguities: each product involves a coordinate and a momentum that commute! In terms of numbered operators

$$\begin{aligned}\hat{L}_1 &\equiv \hat{x}_2\hat{p}_3 - \hat{x}_3\hat{p}_2, \\ \hat{L}_2 &\equiv \hat{x}_3\hat{p}_1 - \hat{x}_1\hat{p}_3, \\ \hat{L}_3 &\equiv \hat{x}_1\hat{p}_2 - \hat{x}_2\hat{p}_1.\end{aligned}\tag{1.7}$$

Note that the angular momentum operators are Hermitian, since \hat{x}_i and \hat{p}_i are and the products can be reordered without cost:

$$\hat{L}_i^\dagger = \hat{L}_i.\tag{1.8}$$

1.1 Quantum mechanical vector identities

We will write triplets of operators as boldfaced vectors, each element of the triplet multiplied by a unit basis vector, just like we do for ordinary vectors. Thus, for example, we have

$$\begin{aligned}\mathbf{r} &\equiv \hat{x}_1\vec{e}_1 + \hat{x}_2\vec{e}_2 + \hat{x}_3\vec{e}_3, \\ \mathbf{p} &\equiv \hat{p}_1\vec{e}_1 + \hat{p}_2\vec{e}_2 + \hat{p}_3\vec{e}_3, \\ \mathbf{L} &\equiv \hat{L}_1\vec{e}_1 + \hat{L}_2\vec{e}_2 + \hat{L}_3\vec{e}_3.\end{aligned}\tag{1.9}$$

These boldface objects are a bit unusual. They are vectors whose components happen to be operators! Moreover, the basis vectors \vec{e}_i must be declared to commute with any of the operators. The boldface objects are useful whenever we want to use the dot products and cross products of three-dimensional space.

Let us, for generality consider vectors \mathbf{a} and \mathbf{b}

$$\begin{aligned}\mathbf{a} &\equiv a_1 \vec{e}_1 + a_2 \vec{e}_2 + a_3 \vec{e}_3, \\ \mathbf{b} &\equiv b_1 \vec{e}_1 + b_2 \vec{e}_2 + b_3 \vec{e}_3,\end{aligned}\tag{1.10}$$

and we will assume that the a_i 's and b_j 's are operators that do not commute. The following are then standard definitions:

$$\begin{aligned}\mathbf{a} \cdot \mathbf{b} &\equiv a_i b_i, \\ (\mathbf{a} \times \mathbf{b})_i &\equiv \epsilon_{ijk} a_j b_k.\end{aligned}\tag{1.11}$$

The order of the operators in the above right-hand sides cannot be changed; it was chosen conveniently, to be the same as the order of the operators on the left-hand sides. We also define,

$$\mathbf{a}^2 \equiv \mathbf{a} \cdot \mathbf{a}.\tag{1.12}$$

Since the operators do not commute, familiar properties of vector analysis do not hold. For example $\mathbf{a} \cdot \mathbf{b}$ is not equal to $\mathbf{b} \cdot \mathbf{a}$. Indeed,

$$\mathbf{a} \cdot \mathbf{b} = a_i b_i = [a_i, b_i] + b_i a_i,\tag{1.13}$$

so that

$$\boxed{\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} + [a_i, b_i].}\tag{1.14}$$

As an application we have

$$\mathbf{r} \cdot \mathbf{p} = \mathbf{p} \cdot \mathbf{r} + [\hat{x}_i, \hat{p}_i],\tag{1.15}$$

The right-most commutator gives $i\hbar \delta_{ii} = 3i\hbar$ so that we have the amusing three-dimensional identity

$$\boxed{\mathbf{r} \cdot \mathbf{p} = \mathbf{p} \cdot \mathbf{r} + 3i\hbar}.\tag{1.16}$$

For cross products we typically have $\mathbf{a} \times \mathbf{b} \neq -\mathbf{b} \times \mathbf{a}$. Indeed,

$$\begin{aligned}(\mathbf{a} \times \mathbf{b})_i &= \epsilon_{ijk} a_j b_k = \epsilon_{ijk} ([a_j, b_k] + b_k a_j) \\ &= -\epsilon_{ikj} b_k a_j + \epsilon_{ijk} [a_j, b_k]\end{aligned}\tag{1.17}$$

where we flipped the k, j indices in one of the epsilon tensors in order to identify a cross product. Indeed, we have now

$$\boxed{(\mathbf{a} \times \mathbf{b})_i = -(\mathbf{b} \times \mathbf{a})_i + \epsilon_{ijk} [a_j, b_k].}\tag{1.18}$$

The simplest example of the use of this identity is one where we use \mathbf{r} and \mathbf{p} . Certainly

$$\mathbf{r} \times \mathbf{r} = 0, \quad \text{and} \quad \mathbf{p} \times \mathbf{p} = 0, \quad (1.19)$$

and more nontrivially,

$$(\mathbf{r} \times \mathbf{p})_i = -(\mathbf{p} \times \mathbf{r})_i + \epsilon_{ijk} [\hat{x}_j, \hat{p}_k]. \quad (1.20)$$

The last term vanishes for it is equal to $i\hbar \epsilon_{ijk} \delta_{jk} = 0$ (the epsilon symbol is antisymmetric in j, k while the delta is symmetric in j, k , resulting in a zero result). We therefore have, quantum mechanically,

$$\boxed{\mathbf{r} \times \mathbf{p} = -\mathbf{p} \times \mathbf{r}.} \quad (1.21)$$

Thus \mathbf{r} and \mathbf{p} can be moved across in the cross product but not in the dot product.

Exercise 1. Prove the following identities for Hermitian conjugation

$$\begin{aligned} (\mathbf{a} \cdot \mathbf{b})^\dagger &= \mathbf{b}^\dagger \cdot \mathbf{a}^\dagger, \\ (\mathbf{a} \times \mathbf{b})^\dagger &= -\mathbf{b}^\dagger \times \mathbf{a}^\dagger. \end{aligned} \quad (1.22)$$

Our definition of the angular momentum operators in (1.7) and the notation developed above imply that we have

$$\boxed{\mathbf{L} = \mathbf{r} \times \mathbf{p} = -\mathbf{p} \times \mathbf{r}.} \quad (1.23)$$

Indeed, given the definition of the product, we have

$$\boxed{\hat{L}_i = \epsilon_{ijk} \hat{x}_j \hat{p}_k.} \quad (1.24)$$

If you write out what this means for $i = 1, 2, 3$ (do it!) you will recover the expressions in (1.7). The angular operator is Hermitian. Indeed, using (1.22) and recalling that \mathbf{r} and \mathbf{p} are Hermitian we have

$$\mathbf{L}^\dagger = (\mathbf{r} \times \mathbf{p})^\dagger = -\mathbf{p}^\dagger \times \mathbf{r}^\dagger = -\mathbf{p} \times \mathbf{r} = \mathbf{L}. \quad (1.25)$$

The use of vector notation implies that, for example,

$$\mathbf{L}^2 = \mathbf{L} \cdot \mathbf{L} = \hat{L}_1 \hat{L}_1 + \hat{L}_2 \hat{L}_2 + \hat{L}_3 \hat{L}_3 = \hat{L}_i \hat{L}_i. \quad (1.26)$$

The classical angular momentum operator is orthogonal to both \vec{r} and \vec{p} as it is built from the cross product of these two vectors. Happily, these properties also hold for the quantum angular momentum. Take for example the dot product of \mathbf{r} with \mathbf{L} to get

$$\mathbf{r} \cdot \mathbf{L} = \hat{x}_i \hat{L}_i = \hat{x}_i \epsilon_{ijk} \hat{x}_j \hat{p}_k = \epsilon_{ijk} \hat{x}_i \hat{x}_j \hat{p}_k = 0. \quad (1.27)$$

The last expression is zero because the \hat{x} 's commute and thus form an object symmetric in i, j , while the epsilon symbol is antisymmetric in i, j . Similarly,

$$\mathbf{p} \cdot \mathbf{L} = \hat{p}_i \hat{L}_i = -\hat{p}_i (\mathbf{p} \times \mathbf{r})_i = -\hat{p}_i \epsilon_{ijk} \hat{p}_j \hat{x}_k = -\epsilon_{ijk} \hat{p}_i \hat{p}_j \hat{x}_k = 0. \quad (1.28)$$

In summary:

$$\boxed{\mathbf{r} \cdot \mathbf{L} = \mathbf{p} \cdot \mathbf{L} = 0.} \quad (1.29)$$

In manipulating vector products the following identity is quite useful

$$\epsilon_{ijk} \epsilon_{ipq} = \delta_{jp} \delta_{kq} - \delta_{jq} \delta_{kp}. \quad (1.30)$$

Its contraction is also needed sometimes:

$$\epsilon_{ijk} \epsilon_{ijq} = 2\delta_{kq}. \quad (1.31)$$

For triple products we find

$$\begin{aligned} [\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_k &= \epsilon_{kji} a_j (\mathbf{b} \times \mathbf{c})_i \\ &= \epsilon_{kji} \epsilon_{ipq} a_j b_p c_q \\ &= -\epsilon_{ijk} \epsilon_{ipq} a_j b_p c_q \\ &= -(\delta_{jp} \delta_{kq} - \delta_{jq} \delta_{kp}) a_j b_p c_q \\ &= a_j b_k c_j - a_j b_j c_k \\ &= [a_j, b_k] c_j + b_k a_j c_j - a_j b_j c_k \\ &= [a_j, b_k] c_j + b_k (\mathbf{a} \cdot \mathbf{c}) - (\mathbf{a} \cdot \mathbf{b}) c_k \end{aligned} \quad (1.32)$$

We can write this as

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b} (\mathbf{a} \cdot \mathbf{c}) - (\mathbf{a} \cdot \mathbf{b}) \mathbf{c} + [a_j, b_k] c_j. \quad (1.33)$$

The first two terms are familiar from vector analysis, the last term is quantum mechanical.

Another familiar relation from vector analysis is the classical

$$(\vec{a} \times \vec{b})^2 \equiv (\vec{a} \times \vec{b}) \cdot (\vec{a} \times \vec{b}) = \vec{a}^2 \vec{b}^2 - (\vec{a} \cdot \vec{b})^2. \quad (1.34)$$

In deriving such an equation you assume that the vector components are commuting numbers, not operators. If we have vector operators additional terms arise.

Exercise 2. Show that

$$\begin{aligned} (\mathbf{a} \times \mathbf{b})^2 &= \mathbf{a}^2 \mathbf{b}^2 - (\mathbf{a} \cdot \mathbf{b})^2 \\ &\quad - a_j [a_j, b_k] b_k + a_j [a_k, b_k] b_j - a_j [a_k, b_j] b_k - a_j a_k [b_k, b_j], \end{aligned} \quad (1.35)$$

and verify that this yields

$$(\mathbf{a} \times \mathbf{b})^2 = \mathbf{a}^2 \mathbf{b}^2 - (\mathbf{a} \cdot \mathbf{b})^2 + \gamma \mathbf{a} \cdot \mathbf{b}, \quad \text{when } [a_i, b_j] = \gamma \delta_{ij}, \quad \gamma \in \mathbb{C}, \quad [b_i, b_j] = 0. \quad (1.36)$$

As an application we calculate \mathbf{L}^2

$$\mathbf{L}^2 = (\mathbf{r} \times \mathbf{p})^2, \quad (1.37)$$

equation (1.36) can be applied with $\mathbf{a} = \mathbf{r}$ and $\mathbf{b} = \mathbf{p}$. Since $[a_i, b_j] = [\hat{x}_i, \hat{p}_j] = i\hbar \delta_{ij}$ we read that $\gamma = i\hbar$, so that

$$\boxed{\mathbf{L}^2 = \mathbf{r}^2 \mathbf{p}^2 - (\mathbf{r} \cdot \mathbf{p})^2 + i\hbar \mathbf{r} \cdot \mathbf{p}.} \quad (1.38)$$

Another useful and simple identity is the following

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}, \quad (1.39)$$

as you should confirm in a one-line computation. In commuting vector analysis this triple product is known to be cyclically symmetric. Note, that in the above no operator has been moved across each other –that's why it holds.

1.2 Properties of angular momentum

A key property of the angular momentum operators is their commutation relations with the \hat{x}_i and \hat{p}_i operators. You should verify that

$$\boxed{\begin{aligned} [\hat{L}_i, \hat{x}_j] &= i\hbar \epsilon_{ijk} \hat{x}_k, \\ [\hat{L}_i, \hat{p}_j] &= i\hbar \epsilon_{ijk} \hat{p}_k. \end{aligned}} \quad (1.40)$$

We say that these equations mean that \mathbf{r} and \mathbf{p} are vectors *under* rotations.

Exercise 3. Use the above relations and (1.18) to show that

$$\mathbf{p} \times \mathbf{L} = -\mathbf{L} \times \mathbf{p} + 2i\hbar \mathbf{p}. \quad (1.41)$$

Hermitization is the process by which we construct a Hermitian operator starting from a non-Hermitian one. Say Ω is not hermitian, its Hermitization Ω_h is defined to be

$$\Omega_h \equiv \frac{1}{2}(\Omega + \Omega^\dagger). \quad (1.42)$$

Exercise 4. Show that the Hermitization of $\mathbf{p} \times \mathbf{L}$ is

$$(\mathbf{p} \times \mathbf{L})_h = \frac{1}{2}(\mathbf{p} \times \mathbf{L} - \mathbf{L} \times \mathbf{p}) = \mathbf{p} \times \mathbf{L} - i\hbar \mathbf{p}. \quad (1.43)$$

Inspired by the behavior of \mathbf{r} and \mathbf{p} under rotations, we declare that an operator \mathbf{u} defined by the triplet $(\hat{u}_1, \hat{u}_2, \hat{u}_3)$ is a *vector under rotations* if

$$[\hat{L}_i, \hat{u}_j] = i\hbar \epsilon_{ijk} \hat{u}_k. \quad (1.44)$$

Exercise 5. Use the above commutator to show that with \mathbf{n} a constant vector,

$$[\mathbf{n} \cdot \mathbf{L}, \mathbf{u}] = -i\hbar \mathbf{n} \times \mathbf{u}. \quad (1.45)$$

Given two operators \mathbf{u} and \mathbf{v} that are vectors under rotations you will show that their dot product is a scalar – it commutes with all \hat{L}_i – and their cross product is a vector:

$$[\hat{L}_i, \mathbf{u} \cdot \mathbf{v}] = 0,$$

$$[\hat{L}_i, (\mathbf{u} \times \mathbf{v})_j] = i\hbar \epsilon_{ijk} (\mathbf{u} \times \mathbf{v})_k.$$

(1.46)

Exercise 6. Prove the above equations.

A number of useful commutator identities follow from (1.46). Most importantly, from the second one, taking $\mathbf{u} = \mathbf{r}$ and $\mathbf{v} = \mathbf{p}$ we get

$$[\hat{L}_i, (\mathbf{r} \times \mathbf{p})_j] = i\hbar \epsilon_{ijk} (\mathbf{r} \times \mathbf{p})_k, \quad (1.47)$$

which gives the celebrated Lie algebra of angular momentum¹

$$[\hat{L}_i, \hat{L}_j] = i\hbar \epsilon_{ijk} \hat{L}_k.$$

(1.48)

More explicitly,

$$\begin{aligned} [\hat{L}_x, \hat{L}_y] &= i\hbar \hat{L}_z, \\ [\hat{L}_y, \hat{L}_z] &= i\hbar \hat{L}_x, \\ [\hat{L}_z, \hat{L}_x] &= i\hbar \hat{L}_y. \end{aligned} \quad (1.49)$$

Note the cyclic nature of these equations: take the first and cycle indices ($x \rightarrow y \rightarrow z \rightarrow x$) to obtain the other two. Another set of examples follows from the first identity in (1.46) using our list of vector operators. For example

$$[\hat{L}_i, \mathbf{r}^2] = [\hat{L}_i, \mathbf{p}^2] = [\hat{L}_i, \mathbf{r} \cdot \mathbf{p}] = 0, \quad (1.50)$$

¹We do not know who celebrated it because we were not invited.

and, very importantly,

$$[\hat{L}_i, \mathbf{L}^2] = 0. \quad (1.51)$$

This equation is the reason the operator \mathbf{L}^2 plays a very important role in the study of central potentials. \mathbf{L}^2 will feature as one of the operators in complete sets of commuting observables. An operator, such as \mathbf{L}^2 , that commutes with all the angular momentum operators is called a “Casimir” of the algebra of angular momentum. Note that the validity of (1.51) just uses the algebra of the \hat{L}_i operators not, for example, how they are built from \mathbf{r} and \mathbf{p} .

Exercise 7. Use (1.18) and the algebra of \hat{L} operators to show that

$$\mathbf{L} \times \mathbf{L} = i\hbar \mathbf{L}. \quad (1.52)$$

This is a very elegant way to express the algebra of angular momentum. In fact, we can show that it is totally equivalent to (1.48). Thus we write

$$\boxed{\mathbf{L} \times \mathbf{L} = i\hbar \mathbf{L} \iff [\hat{L}_i, \hat{L}_j] = i\hbar \epsilon_{ijk} \hat{L}_k.} \quad (1.53)$$

Commutation relations of the form

$$[a_i, b_j] = \epsilon_{ijk} c_k, \quad (1.54)$$

admit a natural rewriting in terms of cross products. From (1.18)

$$(\mathbf{a} \times \mathbf{b})_i + (\mathbf{b} \times \mathbf{a})_i = \epsilon_{ijk} [a_j, b_k] = \epsilon_{ijk} \epsilon_{jkl} c_l = 2 c_i. \quad (1.55)$$

This means that

$$\boxed{[a_i, b_j] = \epsilon_{ijk} c_k \rightarrow \mathbf{a} \times \mathbf{b} + \mathbf{b} \times \mathbf{a} = 2 \mathbf{c}.} \quad (1.56)$$

The arrow does *not* work in the reverse direction. One finds $[a_i, b_j] = \epsilon_{ijk} c_k + s_{ij}$ where $s_{ij} = s_{ji}$ is arbitrary and is *not* determined. If the arrow could be reversed, $\mathbf{a} \times \mathbf{b} + \mathbf{b} \times \mathbf{a} = 0$ would imply that \mathbf{a} and \mathbf{b} commute. We have, however, a familiar example where this does not happen: while $\mathbf{r} \times \mathbf{p} + \mathbf{p} \times \mathbf{r} = 0$ (see (1.23)), the operators \mathbf{r} and \mathbf{p} don’t commute.

For a vector \mathbf{u} under rotations, equation (1.56) becomes

$$\mathbf{L} \times \mathbf{u} + \mathbf{u} \times \mathbf{L} = 2i\hbar \mathbf{u}. \quad (1.57)$$

1.3 The central potential Hamiltonian

Angular momentum plays a crucial role in the study of three-dimensional central potential problems. Those are problems where the Hamiltonian describes a particle moving in a potential $V(r)$ that depends just on r , the distance of the particle to the chosen origin. The Hamiltonian takes the form

$$H = \frac{\mathbf{p}^2}{2m} + V(r). \quad (1.58)$$

When writing the Schrödinger equation in position space we identify

$$\mathbf{p} = \frac{\hbar}{i} \nabla, \quad (1.59)$$

and therefore

$$\mathbf{p}^2 = -\hbar^2 \nabla^2, \quad (1.60)$$

where ∇^2 denotes the Laplacian operator –a second order differential operator. In spherical coordinates the Laplacian is well known and gives us

$$\mathbf{p}^2 = -\hbar^2 \left[\frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{r^2} \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right) \right]. \quad (1.61)$$

Our goal is to relate the “angular” part of the above differential operator to angular momentum operators. This will be done by calculating \mathbf{L}^2 and relating it to \mathbf{p}^2 . Since we had from (1.38)

$$\mathbf{L}^2 = \mathbf{r}^2 \mathbf{p}^2 - (\mathbf{r} \cdot \mathbf{p})^2 + i\hbar \mathbf{r} \cdot \mathbf{p}, \quad (1.62)$$

We solve for \mathbf{p}^2 to get

$$\mathbf{p}^2 = \frac{1}{\mathbf{r}^2} \left[(\mathbf{r} \cdot \mathbf{p})^2 - i\hbar \mathbf{r} \cdot \mathbf{p} + \mathbf{L}^2 \right]. \quad (1.63)$$

Let us now consider the above equation in coordinate space, where \mathbf{p} is a gradient. We then have:

$$\mathbf{r} \cdot \mathbf{p} = \frac{\hbar}{i} r \frac{\partial}{\partial r}, \quad (1.64)$$

and thus

$$(\mathbf{r} \cdot \mathbf{p})^2 - i\hbar \mathbf{r} \cdot \mathbf{p} = -\hbar^2 \left(r \frac{\partial}{\partial r} r \frac{\partial}{\partial r} + r \frac{\partial}{\partial r} \right) = -\hbar^2 \left(r^2 \frac{\partial^2}{\partial r^2} + 2r \frac{\partial}{\partial r} \right). \quad (1.65)$$

It then follows that

$$\frac{1}{\mathbf{r}^2} \left[(\mathbf{r} \cdot \mathbf{p})^2 - i\hbar \mathbf{r} \cdot \mathbf{p} \right] = -\hbar^2 \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right) = -\hbar^2 \frac{1}{r} \frac{\partial^2}{\partial r^2} r, \quad (1.66)$$

where the last step is readily checked by explicit expansion. Back in (1.63) we get

$$\mathbf{p}^2 = -\hbar^2 \frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{r^2} \mathbf{L}^2. \quad (1.67)$$

Comparing with (1.61) we identify \mathbf{L}^2 as the operator

$$\boxed{\mathbf{L}^2 = -\hbar^2 \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right).} \quad (1.68)$$

Note that the units are fully carried by the \hbar^2 in front and that the differential operator is completely angular: it has no radial dependence. Given our expression (1.67) for \mathbf{p}^2 we can now rewrite the three-dimensional Hamiltonian as

$$\boxed{H = \frac{\mathbf{p}^2}{2m} + V(r) = -\frac{\hbar^2}{2m} \frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{2mr^2} \mathbf{L}^2 + V(r).} \quad (1.69)$$

A key property of central potential problems is that the angular momentum operators commute with the Hamiltonian

$$\text{Central potential Hamiltonians: } [\hat{L}_i, H] = 0. \quad (1.70)$$

We have seen that \hat{L}_i commutes with \mathbf{p}^2 so it is only needed to show that the \hat{L}_i commute with $V(r)$. This is eminently reasonable, for \hat{L}_i commutes with $\mathbf{r}^2 = r^2$, so one would expect it to commute with any function of $\sqrt{\mathbf{r}^2} = r$. In the problem set you will consider this question and develop a formal argument that confirms the expectation.

The above commutator implies that the \hat{L}_i operators are conserved in central potentials. Indeed

$$i\hbar \frac{d}{dt} \langle \hat{L}_i \rangle = \langle [\hat{L}_i, H] \rangle = 0. \quad (1.71)$$

We can now consider the issue of complete sets of commuting observables. The list of operators that we have is

$$H, \hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{L}_1, \hat{L}_2, \hat{L}_3, \mathbf{r}^2, \mathbf{p}^2, \mathbf{r} \cdot \mathbf{p}, \mathbf{L}^2, \dots \quad (1.72)$$

where, for the time being, we included all operators up to squares of coordinates, momenta, and angular momenta. Since we want to understand the spectrum of the Hamiltonian, one of the labels of states will be the energy and thus H must be in the list of commuting observables. Because of the potential $V(r)$ none of the \hat{p}_i operators commute the Hamiltonian. Because of the \mathbf{p}^2 term in the Hamiltonian none of the \hat{x}_i commute with the Hamiltonian. Nor will $\mathbf{r}^2, \mathbf{p}^2$ and $\mathbf{r} \cdot \mathbf{p}$. The list is thus reduced to

$$H, \hat{L}_1, \hat{L}_2, \hat{L}_3, \mathbf{L}^2. \quad (1.73)$$

where there are no dots anymore, since without \hat{x}_i or \hat{p}_i there are no other operators to build (recall also that $\mathbf{L} \times \mathbf{L} = i\hbar\mathbf{L}$ and thus it is not new). All the operators in the list commute

with H : the \hat{L}_i as discussed in (1.70), and \mathbf{L}^2 because, after all, it is built from \hat{L}_i . But all the operators do not commute with each other. From the \hat{L}_i we can only pick at most one, for then the other two necessarily do not commute with the chosen one. Happily we can also keep \mathbf{L}^2 because of its Casimir property (1.51). Conventionally, everybody chooses $\hat{L}_3 = \hat{L}_z$ as one element of the set of commuting observables. Thus we have

$$\text{Commuting observables: } H, \hat{L}_z, \mathbf{L}^2. \quad (1.74)$$

We can wonder if this set is complete in the sense that all energy eigenstates are uniquely labelled by the eigenvalues of the above operators. The answer is yes, for the bound state spectrum of a particle that has no other degrees of freedom (say, no spin).

2 Algebraic theory of angular momentum

Hermitian operators $\hat{J}_x, \hat{J}_y, \hat{J}_z$ are said to satisfy the algebra of **angular momentum** if the following commutation relations:

$$[\hat{J}_i, \hat{J}_j] = i\hbar \epsilon_{ijk} \hat{J}_k. \quad (2.1)$$

More explicitly, in components

$$\begin{aligned} [\hat{J}_x, \hat{J}_y] &= i\hbar \hat{J}_z \\ [\hat{J}_y, \hat{J}_z] &= i\hbar \hat{J}_x \\ [\hat{J}_z, \hat{J}_x] &= i\hbar \hat{J}_y. \end{aligned} \quad (2.2)$$

The \hat{J}_i operators could be \hat{L}_i , or \hat{S}_i or something else! Will only use this algebra and the Hermiticity of the operators. From this algebra it also follows that

$$[\hat{J}_i, \mathbf{J}^2] = 0. \quad (2.3)$$

This can be checked explicitly, but our proof of the analogous result (1.51): $[\hat{L}_i, \mathbf{L}^2] = 0$ only used the algebra of the operators \hat{L}_i , so this also holds for the \hat{J}_i , which satisfy the same algebra. It is not convenient to define

$$\begin{aligned} \hat{J}_+ &\equiv \hat{J}_x + i\hat{J}_y, \\ \hat{J}_- &\equiv \hat{J}_x - i\hat{J}_y, \end{aligned} \quad (2.4)$$

such that the two operators are Hermitian conjugates of each other:

$$(\hat{J}_+)^{\dagger} = \hat{J}_-. \quad (2.5)$$

Note that both \hat{J}_x and \hat{J}_y can be solved for in terms of \hat{J}_+ and \hat{J}_- . It is useful to compute the algebra of the operators \hat{J}_+ , \hat{J}_- , and \hat{J}_z . We begin by computing the product $\hat{J}_+ \hat{J}_-$:

$$\hat{J}_+ \hat{J}_- = \hat{J}_x^2 + \hat{J}_y^2 - i[\hat{J}_x, \hat{J}_y] = \hat{J}_x^2 + \hat{J}_y^2 + \hbar \hat{J}_z. \quad (2.6)$$

Together with the product in the opposite order we have

$$\begin{aligned} \hat{J}_+ \hat{J}_- &= \hat{J}_x^2 + \hat{J}_y^2 + \hbar \hat{J}_z, \\ \hat{J}_- \hat{J}_+ &= \hat{J}_x^2 + \hat{J}_y^2 - \hbar \hat{J}_z. \end{aligned} \quad (2.7)$$

From these two we can quickly get the commutator:

$$[\hat{J}_+, \hat{J}_-] = 2\hbar \hat{J}_z. \quad (2.8)$$

Moreover, we obtain two expressions for $\hat{J}_x^2 + \hat{J}_y^2$

$$\hat{J}_x^2 + \hat{J}_y^2 = \hat{J}_+ \hat{J}_- - \hbar \hat{J}_z = \hat{J}_- \hat{J}_+ + \hbar \hat{J}_z. \quad (2.9)$$

Adding \hat{J}_z^2 to both sides of the equation we find

$$\boxed{\mathbf{J}^2 = \hat{J}_+ \hat{J}_- + \hat{J}_z^2 - \hbar \hat{J}_z = \hat{J}_- \hat{J}_+ + \hat{J}_z^2 + \hbar \hat{J}_z.} \quad (2.10)$$

Of course, since \hat{J}_i and \mathbf{J}^2 commute, we also have

$$[\hat{J}_\pm, \mathbf{J}^2] = 0. \quad (2.11)$$

We finally have to compute the commutator of \hat{J}_\pm with \hat{J}_z . This is quickly done:

$$[\hat{J}_z, \hat{J}_+] = [\hat{J}_z, \hat{J}_x] + i[\hat{J}_z, \hat{J}_y] = i\hbar \hat{J}_y + i(-i\hbar \hat{J}_x) = \hbar(\hat{J}_x + i\hat{J}_y) = \hbar \hat{J}_+. \quad (2.12)$$

Similarly, $[\hat{J}_z, \hat{J}_-] = -\hbar \hat{J}_-$ and therefore, all in all

$$\boxed{[\hat{J}_z, \hat{J}_\pm] = \pm \hbar \hat{J}_\pm.} \quad (2.13)$$

This is similar to our harmonic oscillator commutators $[\hat{N}, \hat{a}^\dagger] = \hat{a}^\dagger$ and $[\hat{N}, \hat{a}] = -\hat{a}$, if we identify \hat{N} with \hat{J}_z , \hat{a}^\dagger with \hat{J}_+ and \hat{a} with \hat{J}_- . In the oscillator case we learned from these that, acting on states, \hat{a}^\dagger raises the \hat{N} eigenvalue by one unit while \hat{a} decreases it by one unit. As we will see, \hat{J}_+ adds \hbar to the \hat{J}_z eigenvalue and \hat{J}_- subtracts \hbar to the \hat{J}_z eigenvalue.

Since \mathbf{J}^2 and \hat{J}_z are hermitian and commute, they can be simultaneously diagonalized. In fact, there are no more operators in the angular momentum algebra can be added to this list of simultaneously diagonalizable operators. The common eigenstates form an orthonormal basis

for the relevant vector space. We thus introduce eigenstates $|j, m\rangle$, with $j, m \in \mathbb{R}$, where the first label relates to the \mathbf{J}^2 eigenvalue and the second label to the \hat{J}_z eigenvalue:

$$\begin{aligned}\mathbf{J}^2|j, m\rangle &= \hbar^2 j(j+1) |j, m\rangle, \\ \hat{J}_z |j, m\rangle &= \hbar m |j, m\rangle.\end{aligned}\tag{2.14}$$

The orthonormality of states implies that

$$\langle j', m' | j, m \rangle = \delta_{j', j} \delta_{m', m},\tag{2.15}$$

where we assumed that we will not have to deal with continuous values of j, m that would require delta function normalization. This will be confirmed below. Since j and m are real, the eigenvalues of the hermitian operators are real, as they have to be. The first line shows that the eigenvalue of \mathbf{J}^2 is defined to be $\hbar^2 j(j+1)$. This can seem curious: why not $\hbar^2 j^2$? The answer is convenience, as we will see below. Alternatively, if we know $\hbar^2 j(j+1)$, how do we get j ? For this first note that $\hbar^2 j(j+1)$ must be non-negative:

$$\hbar^2 j(j+1) = \langle j, m | \mathbf{J}^2 | j, m \rangle = \sum_{i=1}^3 \langle j, m | \hat{J}_i \hat{J}_i | j, m \rangle = \sum_{i=1}^3 \| \hat{J}_i | j, m \rangle \|^2 \geq 0,\tag{2.16}$$

where in the first step we used the eigenvalue definition and orthonormality. Therefore the condition

$$j(j+1) \geq 0\tag{2.17}$$

is the only a priori condition on the values of j . Since what matters is the eigenvalue of \mathbf{J}^2 we can use any of the two j 's that give a particular value of $j(j+1)$. As shown in the figure below, the positivity of $j(j+1)$ requires $j \geq 0$ or $j \leq -1$. We can simply use $j \geq 0$.

$$\text{States are labeled as } |j, m\rangle \text{ with } j \geq 0.\tag{2.18}$$

You should not think that there are two different states, with two different j 's associated with the eigenvalue $\hbar j(j+1)$. It is just one state, that we are labeling in an unusual way. Of course, a theory may end up having more than one state with the same \mathbf{J}^2 eigenvalue. In that case we will have more than one state with the same $j > 0$.

Let us now investigate what the operators \hat{J}_{\pm} do when acting on the above eigenstates. Since they commute with \mathbf{J}^2 , the operators J_+ or J_- do not change the j value of a state:

$$\mathbf{J}^2(\hat{J}_{\pm}|j, m\rangle) = \hat{J}_{\pm} \mathbf{J}^2 |j, m\rangle = \hbar j(j+1)(\hat{J}_{\pm}|j, m\rangle),\tag{2.19}$$

so that we must have

$$J_{\pm}|j, m\rangle \propto |j, m'\rangle, \text{ for some } m'.\tag{2.20}$$

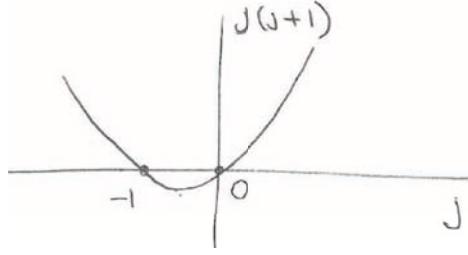


Figure 1: Since $j(j + 1) \geq 0$ for consistency, we can label the states $|j, m\rangle$ using $j \geq 0$.

On the other hand, as anticipated above, the \hat{J}_\pm operators change the value of m :

$$\begin{aligned}\hat{J}_z(\hat{J}_\pm|j, m\rangle) &= ([\hat{J}_z, J_\pm] + J_\pm \hat{J}_z)|j, m\rangle \\ &= (\pm \hbar J_\pm + \hbar m J_\pm)|j, m\rangle \\ &= \hbar(m \pm 1)(J_\pm|j, m\rangle),\end{aligned}\tag{2.21}$$

from which we learn that

$$\hat{J}_\pm|j, m\rangle = C_\pm(j, m)|j, m \pm 1\rangle,\tag{2.22}$$

where $C_\pm(j, m)$ is a constant to be determined. Indeed, we see that \hat{J}_+ raised the m eigenvalue by one unit while \hat{J}_- decreases the m eigenvalue by one unit. To determine $C_\pm(j, m)$ we first take the adjoint of the above equation

$$\langle j, m|\hat{J}_\mp = \langle j, m \pm 1|C_\pm(j, m)^*,\tag{2.23}$$

and then form the overlap

$$\langle j, m|J_\mp J_\pm|j, m\rangle = |C_\pm(j, m)|^2.\tag{2.24}$$

To evaluate the left-hand side use (2.10) in the form

$$\hat{J}_\mp \hat{J}_\pm = \mathbf{J}^2 - \hat{J}_z^2 \mp \hbar \hat{J}_z\tag{2.25}$$

as well as $\langle j, m|j, m\rangle = 1$:

$$|C_\pm(j, m)|^2 = \langle j, m|(\mathbf{J}^2 - \hat{J}_z^2 \mp \hbar \hat{J}_z)|j, m\rangle = \hbar^2 j(j + 1) - \hbar^2 m^2 \mp \hbar^2 m.\tag{2.26}$$

We have thus found that

$$|C_\pm(j, m)|^2 = \hbar^2 (j(j + 1) - m(m \pm 1)) = \|\hat{J}_\pm|j, m\rangle\|^2.\tag{2.27}$$

Here we learn a few things. If we start with a consistent state $|j, m\rangle$ of norm one (as assumed above), the states $\hat{J}_\pm|j, m\rangle \sim |j, m \pm 1\rangle$ created by the action of \hat{J}_\pm on $|j, m\rangle$ are inconsistent if

the middle expression in the above relation is negative. This is because that middle expression is in fact the norm-squared of $\hat{J}_\pm|j, m\rangle$. Assuming that middle expression is positive (or zero) we can take $C_\pm(j, m)$ real and equal to the positive square root

$$C_\pm(j, m) = \hbar\sqrt{j(j+1) - m(m \pm 1)}. \quad (2.28)$$

We have thus obtained

$$J_\pm|j, m\rangle = \hbar\sqrt{j(j+1) - m(m \pm 1)} |j, m \pm 1\rangle.$$

(2.29)

Given a consistent state $|j, m\rangle$, how far can we raise or lower the value of m ? Our classical intuition is that $|\hat{J}_z| \leq |\mathbf{J}|$. So we should get something like $|m| \lesssim \sqrt{j(j+1)}$.

Consider this in two steps:

1. For the raised state to be consistent we must have $\|J_+|j, m\rangle\|^2 \geq 0$ and therefore

$$j(j+1) - m(m+1) \geq 0 \rightarrow m(m+1) \leq j(j+1) \quad (2.30)$$

The solution to the inequality is given in figure 2:

$$-j-1 \leq m \leq j. \quad (2.31)$$

If we had chosen $\hbar^2 j(j+1)$ to be the eigenvalue of \mathbf{J}^2 this inequality would not have had a simple solution.

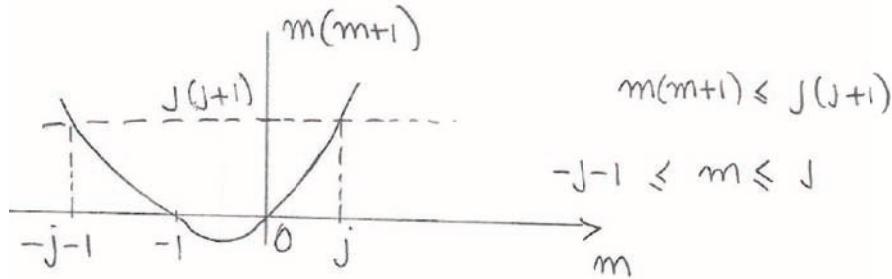


Figure 2: Solving the inequality $m(m+1) \leq j(j+1)$.

Since we are raising m we can focus on the troubles that raising can give given that $m \leq j$. Assume $m = j - \beta$ with $0 < \beta < 1$ so that the inequality (2.31) is satisfied and m is less than one unit below j . Then the raising once gives us a state with $m' = m + 1 > j$ and since the inequality is now violated raising one more time would then give an inconsistent

state. To prevent such inconsistency the process of raising must terminate: there must be a state that raising gives no state (the zero state). That indeed happens only if $m = j$ since then $C_+(j, j) = 0$

$$\hat{J}_+|j, j\rangle = 0. \quad (2.32)$$

2. For the lowered state to be consistent we must have $|\hat{J}_-|j, m\rangle|^2 \geq 0$ and therefore

$$j(j+1) - m(m-1) \geq 0 \rightarrow m(m-1) \leq j(j+1) \quad (2.33)$$

The solution to this inequality is obtained using figure 3 and gives

$$-j \leq m \leq j+1. \quad (2.34)$$

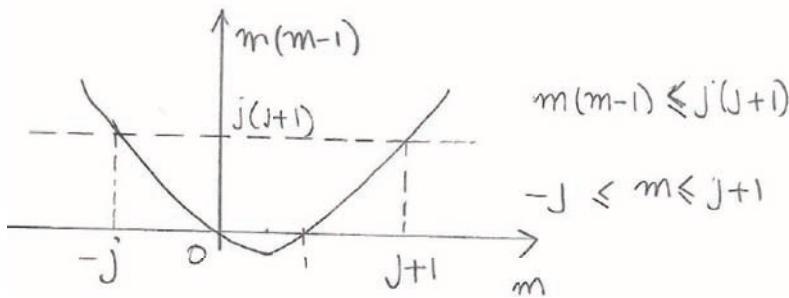


Figure 3: Solving the inequality $m(m-1) \leq j(j+1)$.

This time we can focus here on $m \geq -j$ and the complications due to lowering. Assume $m = -j + \beta$ with $0 < \beta < 1$ so that the constraint (2.34) is satisfied and m is less than one unit above $-j$. Then lowering once gives us a state with $m' = m - 1 < -j$ and since the inequality is now violated lowering one more time would then give an inconsistent state. To prevent such inconsistency we need that lowering terminates on some state for which lowering gives no state (the zero state). That indeed happens only if $m = -j$ since then $C_-(j, -j) = 0$

$$\hat{J}_-|j, -j\rangle = 0. \quad (2.35)$$

The above analysis shows that for consistency a multiplet of states with some given fixed j must be such that the m values must include $-j$ and $+j$. Since m is increased or decreased by integer steps via the \hat{J}_\pm operators, the distance $2j$ between j and $-j$ must be an integer:

$$2j \in \mathbb{Z} \rightarrow j \in \mathbb{Z}/2, \rightarrow j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots \quad (2.36)$$

This is the fundamental quantization of angular momentum. Angular momentum can be integral or half-integral. For any allowed value of j the m values will be $j, j-1, \dots, -j$. Thus the multiplet with angular momentum j has the following $2j+1$ states

$$\begin{aligned} & |j, j\rangle, \\ & |j, j-1\rangle, \\ & \vdots \\ & |j, -j\rangle. \end{aligned} \tag{2.37}$$

For $j = 0$ there is just one state, the **singlet** with $m = 0$: $|0, 0\rangle$.

For $j = \frac{1}{2}$ we have two states:

$$\begin{aligned} & |\frac{1}{2}, \frac{1}{2}\rangle, \\ & |\frac{1}{2}, -\frac{1}{2}\rangle. \end{aligned} \tag{2.38}$$

These are the states of a spin-1/2 particle, when we identify the angular momentum \mathbf{J} with the spin angular momentum operator \mathbf{S} . The top state has $\hat{S}_z = \hbar/2$ and the lower state has $\hat{S}_z = -\hbar/2$. These are our conventional $|+\rangle$ and $|-\rangle$ states, respectively.

For $j = 1$ we have three states :

$$\begin{aligned} & |1, 1\rangle, \\ & |1, 0\rangle, \\ & |1, -1\rangle. \end{aligned} \tag{2.39}$$

For $j = 3/2$ we have four states:

$$\begin{aligned} & |\frac{3}{2}, \frac{3}{2}\rangle \\ & |\frac{3}{2}, \frac{1}{2}\rangle \\ & |\frac{3}{2}, -\frac{1}{2}\rangle \\ & |\frac{3}{2}, -\frac{3}{2}\rangle \end{aligned} \tag{2.40}$$

One last one! For $j = 2$ we have five states:

$$\begin{aligned} & |2, 2\rangle, \\ & |2, 1\rangle, \\ & |2, 0\rangle, \\ & |2, -1\rangle, \\ & |2, -2\rangle. \end{aligned} \tag{2.41}$$

On any state of a multiplet with angular momentum j the eigenvalue J^2 of \mathbf{J}^2 is

$$J^2 = \hbar^2 j(j+1) \rightarrow \frac{1}{\hbar} J = \sqrt{j(j+1)} \tag{2.42}$$

In the limit as j is large

$$\frac{1}{\hbar} J = j \sqrt{1 + \frac{1}{j}} \simeq j + \frac{1}{2} + \mathcal{O}(1/j). \quad (2.43)$$

So for large j the angular momentum is roughly $J \simeq j$.

3 Comments on spherical harmonics

We have constructed the \mathbf{L}^2 operator as a differential operator in position space, with coordinates r, θ, ϕ . The operator happens to depend only on θ and ϕ and takes the form (1.68)

$$\mathcal{L}^2 = -\hbar^2 \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right). \quad (3.1)$$

where we denoted it with a calligraphic symbol to make it clear we are talking about a differential operator. We also have, with the same notation

$$\hat{\mathcal{L}}_z = \frac{\hbar}{i} \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right). \quad (3.2)$$

A short calculation, passing to spherical coordinates (do it!) shows that

$$\hat{\mathcal{L}}_z = \frac{\hbar}{i} \frac{\partial}{\partial \phi}. \quad (3.3)$$

Finally, a longer calculation shows that

$$\mathcal{L}_\pm = \hbar e^{\pm i\phi} \left(i \cot \theta \frac{\partial}{\partial \phi} \pm \frac{\partial}{\partial \theta} \right). \quad (3.4)$$

Recall now how things work for coordinate representations. For a single coordinate x we had that the operator \hat{p} can be taken out of the matrix element as the differential operator \mathfrak{p} :

$$\langle x | \hat{p} | \psi \rangle = \mathfrak{p} \langle x | \psi \rangle, \text{ where } \mathfrak{p} = \frac{\hbar}{i} \frac{\partial}{\partial x} \quad (3.5)$$

We will let $\langle \theta \phi |$ denote position states on the unit sphere and the spherical harmonic $Y_{\ell m}$ will be viewed as the wavefunction for the state $|\ell m\rangle$ so that

$$Y_{\ell m}(\theta, \phi) \equiv \langle \theta \phi | \ell, m \rangle. \quad (3.6)$$

Consider now (2.14) in the form

$$\begin{aligned} \mathbf{L}^2 |\ell, m\rangle &= \hbar^2 \ell(\ell+1) |\ell, m\rangle, \\ \hat{\mathcal{L}}_z |\ell, m\rangle &= \hbar m |\ell, m\rangle. \end{aligned} \quad (3.7)$$

Letting the bra $\langle \theta\phi|$ act on them we have

$$\begin{aligned}\langle \theta\phi | \mathbf{L}^2 | \ell, m \rangle &= \hbar^2 \ell(\ell+1) \langle \theta\phi | \ell, m \rangle, \\ \langle \theta\phi | \hat{L}_z | \ell, m \rangle &= \hbar m \langle \theta\phi | \ell, m \rangle.\end{aligned}\quad (3.8)$$

Using the analog of (3.5) for our operators we have

$$\begin{aligned}\mathcal{L}^2 \langle \theta\phi | \ell, m \rangle &= \hbar^2 \ell(\ell+1) \langle \theta\phi | \ell, m \rangle, \\ \hat{\mathcal{L}}_z \langle \theta\phi | \ell, m \rangle &= \hbar m \langle \theta\phi | \ell, m \rangle.\end{aligned}\quad (3.9)$$

These are equivalent to

$$\begin{aligned}\mathcal{L}^2 Y_{\ell m}(\theta, \phi) &= \hbar^2 \ell(\ell+1) Y_{\ell m}(\theta, \phi), \\ \hat{\mathcal{L}}_z Y_{\ell m}(\theta, \phi) &= \hbar m Y_{\ell m}(\theta, \phi).\end{aligned}\quad (3.10)$$

where \mathcal{L}^2 and $\hat{\mathcal{L}}_z$ are the coordinate representation operators for \mathbf{L}^2 and \hat{L}_z respectively.

On the unit sphere the measure of integration is $\sin \theta d\theta d\phi$ so we postulate that the completeness relation for the $|\theta\phi\rangle$ position states reads

$$\int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\phi |\theta\phi\rangle \langle \theta\phi| = \mathbf{1}. \quad (3.11)$$

The integral will be written more briefly as

$$\int d\Omega |\theta\phi\rangle \langle \theta\phi| = \mathbf{1} \quad (3.12)$$

where

$$\int d\Omega = \int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\phi = - \int_1^{-1} d(\cos \theta) \int_0^{2\pi} d\phi = \int_{-1}^1 d(\cos \theta) \int_0^{2\pi} d\phi. \quad (3.13)$$

Our orthogonality relation

$$\langle \ell', m' | \ell, m \rangle = \delta_{\ell', l} \delta_{m', m}, \quad (3.14)$$

gives, by including a complete set of position states

$$\int d\Omega \langle \ell', m' | \theta\phi \rangle \langle \theta\phi | \ell, m \rangle = \delta_{\ell', l} \delta_{m', m}. \quad (3.15)$$

This gives the familiar orthogonality property of the spherical harmonics:

$$\int d\Omega Y_{\ell' m'}^*(\theta, \phi) Y_{\ell m}(\theta, \phi) = \delta_{\ell', l} \delta_{m', m}. \quad (3.16)$$

Note that the equation

$$\hat{\mathcal{L}}_z Y_{\ell m} = \hbar m Y_{\ell m} \quad (3.17)$$

together with (3.3) implies that

$$Y_{\ell m}(\theta, \phi) = P_{\ell m}(\theta) e^{im\phi}. \quad (3.18)$$

The ϕ dependence of the spherical harmonics is very simple indeed!

One can show that for spherical harmonics, which are related to **orbital** angular momentum, one can only have integer ℓ . While j can be half-integral, any attempt to define spherical harmonics for half-integral ℓ fails. You will indeed show in the homework that this is necessarily the case.

4 The radial equation

Recall that from (1.69) we have

$$H = \frac{\mathbf{p}^2}{2m} + V(r) = -\frac{\hbar^2}{2m} \frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{2mr^2} \mathcal{L}^2 + V(r) \quad (4.1)$$

where we used the differential operator realization \mathcal{L}^2 of the operator \mathbf{L}^2 . The Schrödinger equation will be solved using the following ansatz for energy eigenstates

$$\Psi_{E\ell m}(\mathbf{x}) = f_{E\ell m}(r) Y_{\ell m}(\theta, \phi) \quad (4.2)$$

We have the product of a radial function $f_{E\ell m}(r)$ times an angular function $Y_{\ell m}$ which is an eigenstate of \mathbf{L}^2 and of \hat{L}_z :

$$\mathcal{L}^2 Y_{\ell m} = \hbar^2 \ell(\ell+1) Y_{\ell m}, \quad \hat{L}_z Y_{\ell m} = \hbar m Y_{\ell m} \quad (4.3)$$

Plugging this into the Schrödinger equation $H\Psi = E\Psi$, the $Y_{\ell m}$ dependence can be cancelled out and we get

$$-\frac{\hbar^2}{2m} \frac{1}{r} \frac{d^2}{dr^2} (rf_{E\ell m}) + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} f_{E\ell m} + V(r)f_{E\ell m} = Ef_{E\ell m} \quad (4.4)$$

We note that this equation does not depend on the quantum number m (do not confuse this with the mass m !) Therefore the label m is not needed in the radial function and we let $f_{E\ell m} \rightarrow f_{E\ell}$ so that we have

$$\Psi_{E\ell m}(\mathbf{x}) = f_{E\ell}(r) Y_{\ell m}(\theta, \phi) \quad (4.5)$$

and the differential equation, multiplying by r becomes

$$-\frac{\hbar^2}{2m} \frac{d^2}{dr^2} (rf_{E\ell}) + \left(V(r) + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} \right) (rf_{E\ell}) = E(rf_{E\ell}) \quad (4.6)$$

This suggests writing introducing a modified radial function $u_{E\ell}(r)$ by

$$f_{E\ell}(r) = \frac{u_{E\ell}(r)}{r} \quad (4.7)$$

so that we have

$$\boxed{\Psi_{E\ell m}(\mathbf{x}) = \frac{u_{E\ell}(r)}{r} Y_{\ell m}(\theta, \phi),} \quad (4.8)$$

with radial equation

$$-\frac{\hbar^2}{2m} \frac{d^2 u_{E\ell}}{dr^2} + V_{\text{eff}}(r) u_{E\ell} = E u_{E\ell}, \quad (4.9)$$

where the effective potential V_{eff} constructed by adding to the potential $V(r)$ the centrifugal barrier term proportional to \mathbf{L}^2 :

$$V_{\text{eff}}(r) \equiv V(r) + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} \quad (4.10)$$

This is like a one-dimensional Schrödinger equation in the variable r , but as opposed to our usual problems with $x \in (-\infty, \infty)$, the radius $r \in [0, \infty]$ and we will need some special care for $r = 0$.

The normalization of our wavefunctions proceeds as follows

$$\int d^3\mathbf{x} |\Psi_{E\ell m}(\mathbf{x})|^2 = 1 \quad (4.11)$$

This gives

$$\int r^2 dr d\Omega \frac{|u_{E\ell}(r)|^2}{r^2} Y_{\ell m}^*(\Omega) Y_{\ell m}(\Omega) = 1 \quad (4.12)$$

the angular integral gives one and we get

$$\int_0^\infty dr |u_{E\ell}(r)|^2 = 1 \quad (4.13)$$

a rather natural result for the function $u_{E\ell}$ that plays the role of radial wavefunction.

Behavior of solutions as $r \rightarrow 0$. We claim that

$$\lim_{r \rightarrow 0} u_{E\ell}(r) = 0. \quad (4.14)$$

This requirement does not arise from normalization: as you can see in (4.13) a finite $u_{E\ell}$ at $r = 0$ would cause no trouble. Imagine taking a solution u_{E0} with $\ell = 0$ that approaches a constant as $r \rightarrow 0$:

$$\lim_{r \rightarrow 0} u_{E0}(r) = c \neq 0. \quad (4.15)$$

The full solution Ψ near the origin would then take the form

$$\Psi(\mathbf{x}) \simeq \frac{c}{r} Y_{00} = \frac{c'}{r}, \quad (4.16)$$

since Y_{00} is simply a constant. The problem with this wavefunction is that it simply *does not solve* the Schrödinger equation! You may remember from electromagnetism that the Laplacian of the $1/r$ function has a delta function at the origin, so that as a result

$$\nabla^2 \Psi(\mathbf{x}) = -4\pi c' \delta(\mathbf{x}). \quad (4.17)$$

Since the Laplacian is part of the Hamiltonian, this delta function must be cancelled by some other contribution, but there is none, since the potential $V(r)$ does not have delta functions².

We can learn more about the behavior of the radial solution under the reasonable assumption that the *centrifugal barrier dominates the potential as $r \rightarrow 0$* . In this case the most singular terms of the radial differential equation must cancel each other out, leaving less singular terms that we can ignore in this leading term calculation. So we set:

$$-\frac{\hbar^2}{2m} \frac{d^2 u_{E\ell}}{dr^2} + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} u_{E\ell} = 0, \quad \text{as } r \rightarrow 0. \quad (4.18)$$

or equivalently

$$\frac{d^2 u_{E\ell}}{dr^2} = \frac{\ell(\ell+1)}{r^2} u_{E\ell}, \quad (4.19)$$

The solutions of this can be taken to be $u_{E\ell} = r^s$ with s a constant to be determined. We then find

$$s(s-1) = \ell(\ell+1) \rightarrow s = \ell+1, s = -\ell \quad (4.20)$$

thus leading to two possible behaviors near $r = 0$:

$$u_{E\ell} \sim r^{\ell+1}, \quad u_{E\ell} \sim \frac{1}{r^\ell}. \quad (4.21)$$

For $\ell = 0$ the second behavior was shown to be inconsistent with the Schrödinger equation at $r = 0$ (because of a delta function). For $\ell > 0$ the second behavior is not consistent with normalization. Therefore we have established that

$$u_{E\ell} \sim c r^{\ell+1}, \quad \text{as } r \rightarrow 0.$$

(4.22)

Note that the full radial dependence is obtained by dividing by r , so that

$$f_{E\ell} \sim c r^\ell, \quad (4.23)$$

This allows for a constant non-zero wavefunction at the origin only for $\ell = 0$. Only for $\ell = 0$ a particle can be at the origin. For $\ell \neq 0$ the angular momentum “barrier” prevents the particle from reaching the origin.

²Delta function potentials in more than one dimension are very singular and require regulation.

Behavior of solutions as $r \rightarrow \infty$. Again, we can make some definite statements once we assume some properties of the potential. Let us consider the case when the potential $V(r)$ vanishes beyond some radius or at least decays fast enough as the radius grows without bound

$$V(r) = 0, \text{ for } r > r_0, \text{ or } \lim_{r \rightarrow \infty} rV(r) = 0. \quad (4.24)$$

Curiously, the above assumptions are violated for the $1/r$ potential of the Hydrogen atom (an extended discussion of related issues can be found in Shankar around page 344). Under these assumptions we ignore the effective potential completely (including the centrifugal barrier) and the equation becomes

$$-\frac{\hbar^2}{2m} \frac{d^2 u_{E\ell}}{dr^2} = Eu_{E\ell}(r). \quad (4.25)$$

The equation is the familiar

$$\frac{d^2 u_{E\ell}}{dr^2} = -\frac{2mE}{\hbar^2} u_{E\ell}. \quad (4.26)$$

The resulting $r \rightarrow \infty$ behavior follows immediately

$$\begin{aligned} E < 0, \quad u_{E\ell} &\sim \exp\left(-\sqrt{\frac{2m|E|}{\hbar^2}} r\right), \\ E > 0, \quad u_{E\ell} &\sim \exp(\pm ikr), \quad k = \sqrt{\frac{2mE}{\hbar^2}}. \end{aligned} \quad (4.27)$$

The first behavior, for $E < 0$ is typical of bound states. For $E > 0$ we have a continuous spectrum with degenerate solutions (hence the \pm). Having understood the behavior of solutions near $r = 0$ and for $r \rightarrow \infty$ this allows for qualitative plots of radial solutions.

The discrete spectrum is organized as follows. We have energy eigenstates for all values of ℓ . In fact for each value of ℓ the potential V_{eff} in the radial equation is different. So this equation must be solved for $\ell = 0, 1, \dots$. For each fixed ℓ we have a one-dimensional problem, so we have no degeneracies in the bound state spectrum. We have a set of allowed values of energies that depend on ℓ and are numbered using an integer $n = 1, 2, \dots$. For each allowed energy $E_{n\ell}$ we have a single radial solution $u_{n\ell}$.

$$\text{Fixed } \ell, \text{ Energies: } E_{n\ell}, \quad \text{Radial function: } u_{n\ell}, \quad n = 1, 2, \dots \quad (4.28)$$

Of course each solution $u_{n\ell}$ for the radial equation represents $2\ell + 1$ degenerate solutions to the Schrödinger equation corresponding to the possible values of \hat{L}_z in the range $(-\ell\hbar, \ell\hbar)$. Note that n has replaced the label E in the radial solution, and the energies have now been labeled. This is illustrated in the diagram of Figure 4, where each solution of the radial equation is shown as a short line atop an ℓ label on the horizontal axis. This is the spectral diagram for the central-potential Hamiltonian. Each line of a given ℓ represents the $(2\ell + 1)$ degenerate states obtained with $m = -\ell, \dots, \ell$. Because the bound state spectrum of a one-dimensional potential

is non-degenerate, our radial equation can't have any degeneracies for any fixed ℓ . Thus the lines on the diagram are single lines! Of course, other types of degeneracies of the spectrum can exist: some states having different values of ℓ may have the same energy. In other words, the states may match across columns on the figure. Finally, note that since the potential becomes more positive as ℓ is increased, the lowest energy state occurs for $\ell = 0$ and the energy $E_{1,\ell}$ of the lowest state for a given ℓ increases as we increase ℓ .

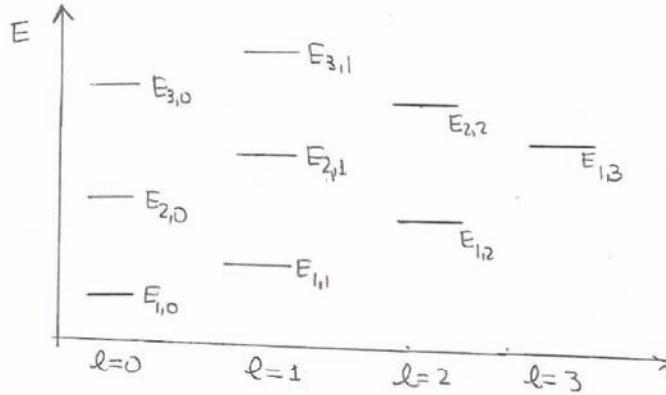


Figure 4: The generic discrete spectrum of a central-potential Hamiltonian, showing the angular momentum ℓ multiplets and their energies.

5 The free particle and the infinite spherical well

5.1 Free particle

It may sound funny at first, but it is interesting to find the radial solutions that correspond to a free particle! A particle that moves in $V(r) = 0$. This amounts to a very different description of the energy eigenstates. In cartesian coordinates we would write solutions as momentum eigenstates, for all values of the momentum. To label such solutions we could use three labels: the components of the momentum. Alternatively, we can use the energy and the direction defined by the momentum, which uses two labels. Here the solutions will be labeled by the energy and (ℓ, m) , the usual two integers that describe the angular dependence (of course, ℓ affects the radial dependence too). The radial equation is

$$-\frac{\hbar^2}{2m} \frac{d^2 u_{E\ell}}{dr^2} + \frac{\hbar^2 \ell(\ell+1)}{2m r^2} u_{E\ell} = Eu_{E\ell} \quad (5.29)$$

which is, equivalently

$$-\frac{d^2 u_{E\ell}}{dr^2} + \frac{\ell(\ell+1)}{r^2} u_{E\ell} = k^2 u_{E\ell}, \quad k \equiv \sqrt{\frac{2mE}{\hbar^2}} \quad (5.30)$$

In this equation there is no quantization of the energy. Indeed we can redefine the radial coordinate in a way that the energy does not appear, namely, k does not appear. Letting $\rho = kr$ the equation becomes

$$-\frac{d^2 u_{E\ell}}{d\rho^2} + \frac{\ell(\ell+1)}{\rho^2} u_{E\ell} = u_{E\ell}, \quad (5.31)$$

The solution of this differential equation with regular behavior at the origin is $u_{E\ell} = c\rho j_\ell(\rho)$, where c is an arbitrary constant. This means that we can take

$$u_{E\ell} = r j_\ell(kr). \quad (5.32)$$

Here the $j_\ell(x)$ are the spherical Bessel functions. All in all we have

$$\text{Free particle: } \Psi_{E\ell m}(\mathbf{x}) = j_\ell(kr) Y_{lm}(\theta, \phi). \quad (5.33)$$

The spherical Bessel functions have the following behavior

$$x j_\ell(x) \sim \frac{x^{\ell+1}}{(2\ell+1)!!}, \text{ as } x \rightarrow 0, \quad \text{and} \quad x j_\ell(x) \sim \sin\left(x - \frac{\ell\pi}{2}\right) \text{ as } x \rightarrow \infty. \quad (5.34)$$

which implies the correct behavior for $u_{E\ell}$ as $r \rightarrow 0$ and $r \rightarrow \infty$. Indeed, for $r \rightarrow \infty$ we have

$$\text{Free particle : } u_{E\ell} \sim \sin\left(kr - \frac{\ell\pi}{2}\right), \text{ as } r \rightarrow \infty. \quad (5.35)$$

Whenever the potential is not zero, but vanishes beyond some radius, the solutions, for $r \rightarrow \infty$ take the form

$$u_{E\ell} \sim \sin\left(kr - \frac{\ell\pi}{2} + \delta_\ell(E)\right), \text{ as } r \rightarrow \infty. \quad (5.36)$$

Here $\delta_\ell(E)$ is called the **phase shift** and by definition vanishes if there is no potential. The form of the solution above is consistent with our general behavior, as this is a superposition of the two solutions available in (4.27) for $E > 0$. The phase shift contains all the information about a potential $V(r)$ available to someone probing the potential from far away by scattering particles off of it.

5.2 The infinite spherical well

An infinite spherical well of radius a is a potential that forces the particle to be within the sphere $r = a$. The potential is zero for $r \leq a$ and it is infinite for $r > a$.

$$V(r) = \begin{cases} 0, & \text{if } r \leq a \\ \infty, & \text{if } r > a \end{cases} \quad (5.37)$$

The Schrödinger radial equation is the same as the one for the free particle

$$-\frac{d^2 u_{E\ell}}{d\rho^2} + \frac{\ell(\ell+1)}{\rho^2} u_{E\ell} = u_{E\ell}, \quad \rho = kr \quad (5.38)$$

where k again encodes the energy E , which is greater than zero. It follows that the solutions are the ones we had before, with spherical Bessel functions, but this time quantization of the energy arises because the wavefunctions must vanish for $r = a$.

Let us do the case $\ell = 0$ without resorting to the Bessel functions. The above equation becomes

$$-\frac{d^2 u_{E,0}}{d\rho^2} = u_{E,0} \quad \rightarrow \quad u_{E,0} = A \sin \rho + B \cos \rho. \quad (5.39)$$

Since the solution must vanish at $r = 0$ we must choose the sin function:

$$u_{E,0}(r) = \sin kr. \quad (5.40)$$

Since this must vanish for $r = a$ we have that k must take values k_n with

$$k_n a = n\pi, \quad \text{for } n = 1, 2, \dots \infty. \quad (5.41)$$

Those values of k_n correspond to energies $E_{n,0}$ where the n indexes the solutions and the 0 represents $\ell = 0$:

$$E_{n,0} = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2}{2ma^2} (k_n a)^2 = \frac{\hbar^2}{2ma^2} n^2 \pi^2. \quad (5.42)$$

Note that $\frac{\hbar^2}{2ma^2}$ is the natural energy scale for this problem and therefore it is convenient to define the unit-free scaled energies $\mathcal{E}_{n,\ell}$ by dividing $E_{n,\ell}$ by the natural energy by

$$\mathcal{E}_{n,\ell} \equiv \frac{2ma^2}{\hbar^2} E_{n,\ell}. \quad (5.43)$$

It follows that the ‘energies’ for $\ell = 0$ are

$$\mathcal{E}_{n,0} = n^2 \pi^2, \quad u_{n,0} = \sin\left(\frac{n\pi r}{a}\right). \quad (5.44)$$

We have

$$\mathcal{E}_{1,0} \simeq 9.8696, \quad \mathcal{E}_{2,0} \simeq 39.478, \quad \mathcal{E}_{3,0} \simeq 88.826, \quad (5.45)$$

Let us now do $\ell = 1$. Here the solutions are $u_{E,1} = r j_1(kr)$. This Bessel function is

$$j_1(\rho) = \frac{\sin \rho}{\rho^2} - \frac{\cos \rho}{\rho} \quad (5.46)$$

The zeroes of $j_1(\rho)$ occur for $\tan \rho = \rho$. Of course, we are not interested in the zero at $\rho = 0$. You can check that the first three zeroes occur for 4.4934, 7.7252, 10.904. For higher values of ℓ it becomes a bit more complicated but there are tables of zeroes on the web.

There is notation in which the nontrivial zeroes are denoted by $z_{n,\ell}$ where

$$z_{n,\ell} \text{ is the } n\text{-th zero of } j_\ell : j_\ell(z_{n,\ell}) = 0. \quad (5.47)$$

The vanishing condition at $r = a$ quantizes k so that

$$k_{n,\ell} a = z_{n,\ell} \quad (5.48)$$

and the energies

$$E_{n,\ell} = \frac{\hbar^2}{2ma^2} (k_{n,\ell}a)^2 \rightarrow \mathcal{E}_{n,\ell} = z_{n,\ell}^2 \quad (5.49)$$

We have

$$\begin{aligned} z_{1,1} &= 4.4934, & z_{2,1} &= 7.7252, & z_{3,1} &= 10.904 \\ z_{1,2} &= 5.7634, & z_{2,2} &= 9.095, \\ z_{1,3} &= 6.9879, & z_{2,3} &= 10.417. \end{aligned} \quad (5.50)$$

which give us

$$\begin{aligned} \mathcal{E}_{1,1} &= 20.191, & \mathcal{E}_{2,1} &= 59.679, & \mathcal{E}_{3,1} &= 118.89 \\ \mathcal{E}_{1,2} &= 33.217, & \mathcal{E}_{2,2} &= 82.719, \\ \mathcal{E}_{1,3} &= 48.83, & \mathcal{E}_{2,3} &= 108.51. \end{aligned} \quad (5.51)$$

The main point to be made is that there are no accidental degeneracies: the energies for different values of ℓ never coincide. More explicitly, with $\ell \neq \ell'$ we have that $\mathcal{E}_{n,\ell} \neq \mathcal{E}_{n',\ell'}$ for any choices of n and n' . This is illustrated in figure 5.

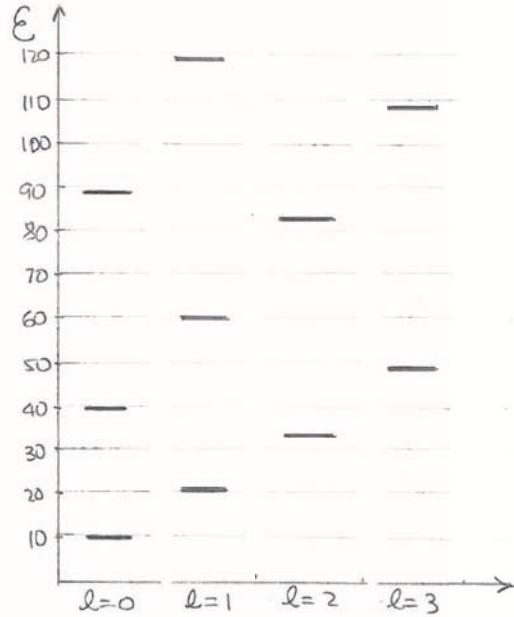


Figure 5: The spectrum of the infinite spherical square well. There are no accidental degeneracies.

6 The three-dimensional isotropic oscillator

The potential of the 3D isotropic harmonic oscillator is as follows:

$$V = \frac{1}{2}m\omega^2(x^2 + y^2 + z^2) = \frac{1}{2}m\omega^2r^2. \quad (6.52)$$

As we will see, the spectrum for this quantum mechanical system has degeneracies, that are explained by the existence of some **hidden symmetry**, a symmetry that is not obvious from the start. Thus in some ways this quantum 3D oscillator is a lot more symmetric than the infinite spherical well.

As you know, for the 3D oscillator we can use creation and annihilation operators $\hat{a}_x^\dagger, \hat{a}_y^\dagger, \hat{a}_z^\dagger$ and $\hat{a}_x, \hat{a}_y, \hat{a}_z$ associated with 1D oscillators in the x, y , and z directions. The Hamiltonian then takes the form:

$$H = \hbar\omega(\hat{N}_1 + \hat{N}_2 + \hat{N}_3 + \frac{3}{2}) = \hbar\omega(\hat{N} + \frac{3}{2}). \quad (6.53)$$

where we defined $\hat{N} \equiv \hat{N}_1 + \hat{N}_2 + \hat{N}_3$.

We now want to explain how tensor products are relevant to the 3D oscillator. We have discussed tensor products before to describe two particles, each associated with a vector space and the combined system associated with the tensor product of vector spaces. But tensor products are also relevant to single particles, if they have degrees of freedom that live in different spaces, or more than one set of attributes, each of which described by states in some vector space. For example, if a spin 1/2 particle can move, the relevant states live in the tensor product of momentum space and the 2-dimensional complex vector space of spin. States are obtained by superposition of basic states of the form $|p\rangle \otimes (\alpha|+\rangle + \beta|-\rangle)$

For the 3D oscillator, the Hamiltonian is the sum of commuting Hamiltonians of 1D oscillators for the x, y , and z directions. Thus the general states are obtained by tensoring the state spaces $\mathcal{H}_x, \mathcal{H}_y$, and \mathcal{H}_z of the three independent oscillators. It is a single particle oscillating, but the description of what it is doing entails saying what is doing in each of the independent directions. Thus we write

$$\mathcal{H}_{3D} = \mathcal{H}_x \otimes \mathcal{H}_y \otimes \mathcal{H}_z. \quad (6.54)$$

Instead of this tensor product reflecting the behavior of three different particles, this tensor product allows us to describe the behavior of one particle in three different directions. The vacuum state $|0\rangle$ of the 3D oscillator can be viewed as

$$|0\rangle \equiv |0\rangle_x \otimes |0\rangle_y \otimes |0\rangle_z. \quad (6.55)$$

The associated wavefunction is

$$\Psi(x, y, z) = \langle x| \otimes \langle y| \otimes \langle z| |0\rangle = \langle x|0\rangle_x \langle y|0\rangle_y \langle z|0\rangle_z = \psi_0(x)\psi_0(y)\psi_0(z). \quad (6.56)$$

where ψ_0 is the ground state wavefunction of the 1D oscillator. This is the expected answer. Recalling the form of (non-normalized) basis states for \mathcal{H}_x , \mathcal{H}_y , and \mathcal{H}_z :

$$\begin{aligned} \text{basis states for } \mathcal{H}_x : & (\hat{a}_x^\dagger)^{n_x} |0\rangle_x, n_x = 0, 1, \dots \\ \text{basis states for } \mathcal{H}_y : & (\hat{a}_y^\dagger)^{n_y} |0\rangle_y, n_y = 0, 1, \dots \\ \text{basis states for } \mathcal{H}_z : & (\hat{a}_z^\dagger)^{n_z} |0\rangle_z, n_z = 0, 1, \dots \end{aligned} \quad (6.57)$$

We then have that the basis states for the 3D state space are

$$\text{basis states of } \mathcal{H}_{3D} : (\hat{a}_x^\dagger)^{n_x} |0\rangle_x \otimes (\hat{a}_y^\dagger)^{n_y} |0\rangle_y \otimes (\hat{a}_z^\dagger)^{n_z} |0\rangle_z, n_x, n_y, n_z \in \{0, 1, \dots\} \quad (6.58)$$

This is what we would expect intuitively, we simply pile arbitrary numbers of \hat{a}_x^\dagger , \hat{a}_y^\dagger , and \hat{a}_z^\dagger on the vacuum. It is this multiplicative structure that is the signature of tensor products. Having understood the above, for brevity we write such basis states simply as

$$(\hat{a}_x^\dagger)^{n_x} (\hat{a}_y^\dagger)^{n_y} (\hat{a}_z^\dagger)^{n_z} |0\rangle. \quad (6.59)$$

Each of the states in (6.58) has a wavefunction that is the product of x , y , and z -dependent wavefunctions. Once we form superpositions of such states, the total wavefunction cannot any longer be factorized into x , y , and z -dependent wavefunctions. The x , y , and z -dependences become ‘entangled’. These are precisely the analogs of entangled states of three particles.

We are ready to begin constructing the individual states of the 3D isotropic harmonic oscillator system. The key property is that the states must organize themselves into representations of angular momentum. Since angular momentum commutes with the Hamiltonian, angular momentum multiplets represent degenerate states.

We already built the ground state, which is a single state with \hat{N} eigenvalue $N = 0$. All other states have higher energies, so this state must be, by itself a representation of angular momentum. It can only be the singlet $\ell = 0$. Thus we have

$$N = 0, E = \frac{3}{2}\hbar\omega, |0\rangle \leftrightarrow \ell = 0. \quad (6.60)$$

The states with $N = 1$ have $E = (5/2)\hbar\omega$ and are

$$\hat{a}_x^\dagger |0\rangle, \hat{a}_y^\dagger |0\rangle, \hat{a}_z^\dagger |0\rangle \quad (6.61)$$

These three states fit precisely into an $\ell = 1$ multiplet (a triplet). There is no other possibility, in fact: any higher ℓ multiplet has too many states and we only have 3 degenerate ones. Moreover, we cannot have three singlets, this is a degeneracy inconsistent with the lack of degeneracy for 1D bound states (as discussed earlier). The $\ell = 0$ ground state and the $\ell = 1$ triplet at the first excited level are indicated in Figure 7.

Let us proceed now with the states at $N = 2$ or $E = (7/2)\hbar\omega$. These are, the following six states:

$$(\hat{a}_x^\dagger)^2|0\rangle, (\hat{a}_y^\dagger)^2|0\rangle, (\hat{a}_z^\dagger)^2|0\rangle, \hat{a}_x^\dagger\hat{a}_y^\dagger|0\rangle, \hat{a}_x^\dagger\hat{a}_z^\dagger|0\rangle, \hat{a}_y^\dagger\hat{a}_z^\dagger|0\rangle. \quad (6.62)$$

To help ourselves in trying to find the angular momentum multiplets recall that that the number of states $\#$ for a given ℓ are

ℓ	$\#$
0	1
1	3
2	5
3	7
4	9
5	11
6	13
7	15

Since we cannot use the triplet twice, the only way to get six states is having five from $\ell = 2$ and one from $\ell = 0$. Thus

$$\text{Six } N = 2 \text{ states : } (\ell = 2) \oplus (\ell = 0). \quad (6.63)$$

Note that here we use the direct sum (not the tensor product!) the six states define a six dimensional vector space spanned by five vectors in $\ell = 0$ and one vector in $\ell = 0$. Had we used a tensor product we would just have 5 vectors.

Let us continue to figure out the pattern. At $N = 3$ with $E = (9/2)\hbar\omega$ we actually have 10 states (count them!) It would seem now that there are two possibilities for multiplets

$$(\ell = 3) \oplus (\ell = 1) \text{ or } (\ell = 4) \oplus (\ell = 0) \quad (6.64)$$

We can argue that the second possibility cannot be. The problem with it is that the $\ell = 3$ multiplet, which has not appeared yet, would not arise at this level. If it would arise later, it would do so at a higher energy, and we would have the lowest $\ell = 3$ multiplet above the lowest $\ell = 4$ multiplet, which is not possible. You may think that perhaps $\ell = 3$ multiplets never appear and the inconsistency is avoided, but this is not true. At any rate we will give below a more rigorous argument. The conclusion, however is that

$$\text{Ten } N = 3 \text{ states : } (\ell = 3) \oplus (\ell = 1). \quad (6.65)$$

Let us do the next level! At $N = 4$ we find 15 states. Instead of writing them out let us count them without listing them. In fact, we can easily do the general case of arbitrary integer $N \geq 1$. The states we are looking for are of the form

$$(\hat{a}_x^\dagger)^{n_x}(\hat{a}_y^\dagger)^{n_y}(\hat{a}_z^\dagger)^{n_z}|0\rangle, \text{ with } n_x + n_y + n_z = N. \quad (6.66)$$

We need to count how many different solutions there are to $n_x + n_y + n_z = N$, with $n_x, n_y, n_z \geq 0$. This is the number of states $\#(N)$ at level N . To visualize this think of $n_x + n_y + n_z = N$ as the equation for a plane in three-dimensional space with axes n_x, n_y, n_z . Since no integer can be negative, we are looking for points with integer coordinates in the region of the plane that lies on the positive octant, as shown in Figure 6. Starting at one of the three corners, say $(n_x, n_y, n_z) = (N, 0, 0)$ we have one point, then moving towards the origin we encounter two points, then three, and so on until we find $N+1$ points on the (n_y, n_z) plane. Thus, the number of states $\#(N)$ for number N is

$$\#(N) = 1 + 2 + \dots + (N+1) = \frac{(N+1)(N+2)}{2} \quad (6.67)$$

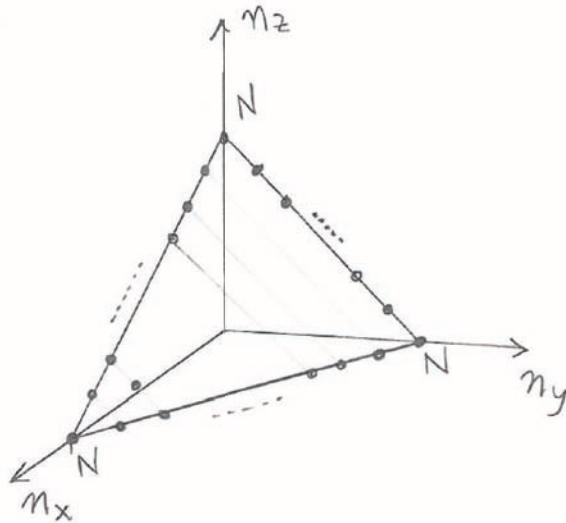


Figure 6: Counting the number of degenerate states with number N in the 3D simple harmonic oscillator.

Back to the $N = 4$ level, $\#(4)=15$. We rule out a single $\ell = 7$ multiplet since states with $\ell = 4, 5, 6$ have not appeared yet. By this logic the highest ℓ multiplet for $N = 4$ must be the lowest that has not appeared yet, thus $\ell = 4$, with 9 states. The remaining six must appear as $\ell = 2$ plus $\ell = 0$. Thus, we have

$$15 \text{ } N = 4 \text{ states : } (\ell = 4) \oplus (\ell = 2) \oplus (\ell = 0) . \quad (6.68)$$

Thus we see that ℓ jumps by steps of two, starting from the maximal ℓ . This is in fact the rule. It is quickly confirmed for the $\#(5)=21$ states with $N = 5$ would arise from $(\ell = 5) \oplus (\ell = 3) \oplus (\ell = 1)$. All this is shown in Figure 7.

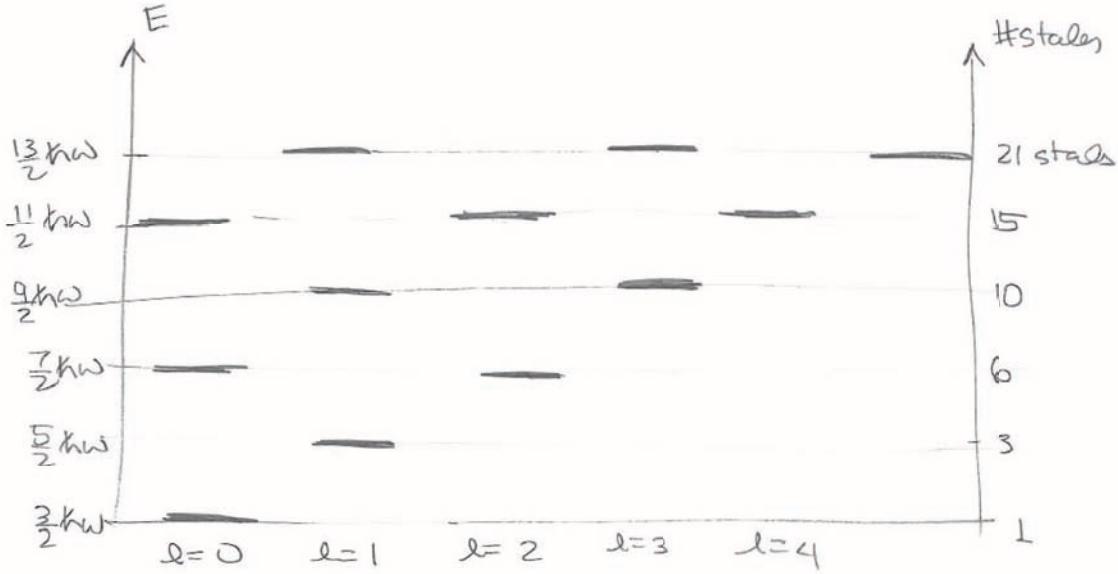


Figure 7: Spectral diagram for angular momentum multiplets in the 3D isotropic harmonic oscillator.

Some of the structure of angular momentum multiplets can be seen more explicitly using the \hat{a}_L and \hat{a}_R operators introduced for the 2D harmonic oscillator:

$$\hat{a}_L = \frac{1}{\sqrt{2}}(\hat{a}_x + i\hat{a}_y), \quad \hat{a}_R = \frac{1}{\sqrt{2}}(\hat{a}_x - i\hat{a}_y). \quad (6.69)$$

L and R objects commute with each other and we have $[\hat{a}_L, \hat{a}_L^\dagger] = [\hat{a}_R, \hat{a}_R^\dagger] = 1$. With number operators $\hat{N}_R = \hat{a}_R^\dagger \hat{a}_R$ and $\hat{N}_L = \hat{a}_L^\dagger \hat{a}_L$ we then have $H = \hbar\omega(\hat{N}_R + \hat{N}_L + \hat{N}_z + \frac{3}{2})$ and, more importantly, the z component \hat{L}_z of angular momentum takes the simple form

$$\hat{L}_z = \hbar(\hat{N}_R - \hat{N}_L). \quad (6.70)$$

Note that \hat{a}_z carries no z -component of angular momentum. States are now built acting with arbitrary numbers of \hat{a}_L^\dagger , \hat{a}_R^\dagger and \hat{a}_z^\dagger operators on the vacuum. The $N = 1$ states are then presented as

$$\hat{a}_R^\dagger |0\rangle, \quad \hat{a}_z^\dagger |0\rangle, \quad \hat{a}_L^\dagger |0\rangle. \quad (6.71)$$

We see that the first state has $L_z = \hbar$, the second $L_z = 0$ and the third $L_z = -\hbar$, exactly the three expected values of the $\ell = 1$ multiplet identified before. For number $N = 2$ the state with highest L_z is $(\hat{a}_R^\dagger)^2 |0\rangle$ and it has $L_z = 2\hbar$. This shows that the highest ℓ multiplet is $\ell = 2$. For arbitrary positive integer number N , the state with highest L_z is $(\hat{a}_R^\dagger)^N |0\rangle$ and it has $L_z = \hbar N$. This shows we must have an $\ell = N$ multiplet. This is in fact what we got before! We can also understand the reason for the jump of two units from the top state of the multiplet. Consider

the above state with maximal \hat{L}_z/\hbar equal to N and then the states with one and two units less of \hat{L}_z/\hbar :

$$\begin{aligned}\hat{L}_z/\hbar = N &: (\hat{a}_R^\dagger)^N |0\rangle \\ \hat{L}_z/\hbar = N - 1 &: (\hat{a}_R^\dagger)^{N-1} \hat{a}_z^\dagger |0\rangle \\ \hat{L}_z/\hbar = N - 2 &: (\hat{a}_R^\dagger)^{N-2} (\hat{a}_z^\dagger)^2 |0\rangle, \quad (\hat{a}_R^\dagger)^{N-1} \hat{a}_L^\dagger |0\rangle\end{aligned}\tag{6.72}$$

While there is only one state with one unit less of \hat{L}_z/\hbar there are two states with two units less. One linear combination of these two states must belong to the $\ell = N$ multiplet, but the other linear combination must be the top state of an $\ell = N - 2$ multiplet! This is the reason for the jump of two units.

For arbitrary N we can see why $\#(N)$ can be reproduced by ℓ multiplets skipping by two

$$\begin{aligned}N \text{ odd : } \#(N) &= \underbrace{1 + 2}_{\ell=1} + \underbrace{3 + 4}_{\ell=3} + \underbrace{5 + 6}_{\ell=5} + \underbrace{7 + 8}_{\ell=7} + \dots + \underbrace{N + (N+1)}_{\ell=N} \\ N \text{ even : } \#(N) &= \underbrace{1}_{\ell=0} + \underbrace{2 + 3}_{\ell=2} + \underbrace{4 + 5}_{\ell=4} + \underbrace{6 + 7}_{\ell=6} + \dots + \underbrace{N + (N+1)}_{\ell=N}\end{aligned}\tag{6.73}$$

The accidental degeneracy is “explained” if we identify an operator that commutes with the Hamiltonian (a symmetry) and connects the various ℓ multiplets that appear for a fixed number N . One such operator is

$$K \equiv \hat{a}_R^\dagger \hat{a}_L .\tag{6.74}$$

You can check it commutes with the Hamiltonian, and with a bit more work, that acting on the top state of the $\ell = N - 2$ multiplet it gives the top state of the $\ell = N$ multiplet.

7 Hydrogen atom and Runge-Lenz vector

The hydrogen atom Hamiltonian is

$$H = \frac{\mathbf{p}^2}{2m} - \frac{e^2}{r} .\tag{7.75}$$

The natural length scale here is the Bohr radius a_0 , which is the unique length that can be built using the constants in this Hamiltonian: \hbar , m , and e^2 . We determine a_0 by setting $p \sim \hbar/a_0$ and equating magnitudes of kinetic and potential terms, ignoring numerical factors:

$$\frac{\hbar^2}{ma_0^2} = \frac{e^2}{a_0} \rightarrow a_0 = \frac{\hbar^2}{me^2} \simeq 0.529 \text{ \AA} .\tag{7.76}$$

Note that if the charge of the electron e^2 is decreased, the attraction force decreases and, correctly, the Bohr radius increases. The Bohr radius is the length scale of the hydrogen atom.

A natural energy scale E_0 is

$$E_0 = \frac{e^2}{a_0} = \frac{e^4 m}{\hbar^2} = \left(\frac{e^2}{\hbar c}\right)^2 mc^2 = \alpha^2 (mc^2) \quad (7.77)$$

where we see the appearance of the fine-structure constant α that, in cgs units, takes the form

$$\alpha \equiv \frac{e^2}{\hbar c} \simeq \frac{1}{137}. \quad (7.78)$$

We thus see that the natural energy scale of the hydrogen atom is about $\alpha^2 \simeq 1/18770$ smaller than the rest energy of the electron. This gives about $E_0 = 27.2\text{eV}$. In fact $-E_0/2 = -13.6\text{eV}$ is the bound state energy of the electron in the ground state of the hydrogen atom.

One curious way to approach the calculation of the ground state energy and ground state wavefunction is to factorize the Hamiltonian. One can show that

$$H = \gamma + \frac{1}{2m} \sum_{k=1}^3 \left(\hat{p}_k + i\beta \frac{\hat{x}_k}{r} \right) \left(\hat{p}_k - i\beta \frac{\hat{x}_k}{r} \right) \quad (7.79)$$

for suitable constants β and γ that you can calculate. The ground state $|\Psi_0\rangle$ is then the state for which

$$\left(\hat{p}_k - i\beta \frac{\hat{x}_k}{r} \right) |\Psi_0\rangle = 0. \quad (7.80)$$

The spectrum of the hydrogen atom is described in Figure 8. The energy levels are $E_{\nu\ell}$, where we used $\nu = 1, 2, \dots$, instead of n to label the various solutions for a given ℓ . This is because the label n is reserved for what is called the “principal quantum number”. The degeneracy of the system is such that multiplets with equal $n \equiv \nu + \ell$ have the same energy, as you can see in the figure. Thus, for example, $E_{2,0} = E_{1,1}$, which is to say that the first excited solution for $\ell = 0$ has the same energy as the lowest energy solution for $\ell = 1$. It is also important to note that for any fixed value of n the allowed values of ℓ are

$$\ell = 0, 1, \dots, n-1 \quad (7.81)$$

Finally, the energies are given by

$$E_{\nu\ell} = -\frac{e^2}{2a_0} \frac{1}{(\nu + \ell)^2}, \quad n \equiv \nu + \ell. \quad (7.82)$$

The large amount of degeneracy in this spectrum asks for an explanation. The hydrogen Hamiltonian has in fact some hidden symmetry. It has to do with the so-called Runge-Lenz vector. In the following we discuss the classical origin of this conserved vector quantity.

Imagine we have an energy functional

$$E = \frac{\mathbf{p}^2}{2m} + V(r) \quad (7.83)$$

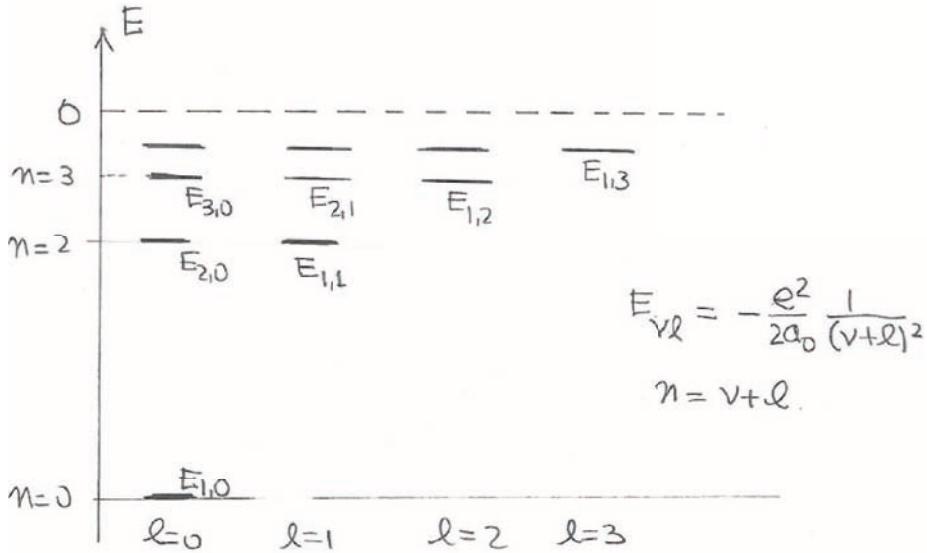


Figure 8: Spectrum of angular momentum multiplets for the hydrogen atom. Here $E_{\nu\ell}$ with $\nu = 1, 2, \dots$, denotes the energy of the ν -th solution for any fixed ℓ . States with equal values of $n \equiv \nu + \ell$ are degenerate. For any fixed n , the values of ℓ run from zero to $n - 1$. Correction: the $n = 0$ in the figure should be $n = 1$.

then the force on the particle moving in this potential is

$$\mathbf{F} = -\nabla V = -V'(r) \frac{\mathbf{r}}{r}, \quad (7.84)$$

where primes denote derivatives with respect to the argument. Newton's equation is

$$\frac{d\mathbf{p}}{dt} = -V'(r) \frac{\mathbf{r}}{r} \quad (7.85)$$

and it is simple to show (do it!) that in this central potential the angular momentum is conserved

$$\frac{d\mathbf{L}}{dt} = 0. \quad (7.86)$$

We now calculate (all classically) the time derivative of $\mathbf{p} \times \mathbf{L}$:

$$\begin{aligned} \frac{d}{dt}(\mathbf{p} \times \mathbf{L}) &= \frac{d\mathbf{p}}{dt} \times \mathbf{L} = -\frac{V'(r)}{r} \mathbf{r} \times (\mathbf{r} \times \mathbf{p}) \\ &= -\frac{mV'(r)}{r} \mathbf{r} \times (\mathbf{r} \times \dot{\mathbf{r}}) \\ &= -\frac{mV'(r)}{r} [\mathbf{r}(\mathbf{r} \cdot \dot{\mathbf{r}}) - \dot{\mathbf{r}} r^2]. \end{aligned} \quad (7.87)$$

We now note that

$$\mathbf{r} \cdot \dot{\mathbf{r}} = \frac{1}{2} \frac{d}{dt}(\mathbf{r} \cdot \mathbf{r}) = \frac{1}{2} \frac{d}{dt} r^2 = r \dot{r}. \quad (7.88)$$

Using this

$$\begin{aligned}\frac{d}{dt}(\mathbf{p} \times \mathbf{L}) &= -\frac{mV'(r)}{r} [\mathbf{r} r\dot{r} - \dot{\mathbf{r}} r^2] = mV'(r)r^2 \left[\frac{\dot{\mathbf{r}}}{r} - \frac{\mathbf{r} \dot{r}}{r^2} \right] \\ &= mV'(r)r^2 \frac{d}{dt} \left(\frac{\mathbf{r}}{r} \right)\end{aligned}\quad (7.89)$$

Because of the factor $V'(r)r^2$, the right-hand side fails to be a total time derivative. But if we focus on potentials for which this factor is a constant we will get a conservation law. So, assume

$$V'(r)r^2 = \gamma, \quad (7.90)$$

for some constant γ . Then

$$\frac{d}{dt}(\mathbf{p} \times \mathbf{L}) = m\gamma \frac{d}{dt} \left(\frac{\mathbf{r}}{r} \right) \rightarrow \frac{d}{dt} \left(\mathbf{p} \times \mathbf{L} - m\gamma \frac{\mathbf{r}}{r} \right) = 0 \quad (7.91)$$

We got a conservation law: that complicated vector inside the parenthesis is constant in time! Back to (7.90) we have

$$\frac{dV}{dr} = \frac{\gamma}{r^2} \rightarrow V(r) = -\frac{\gamma}{r} + c_0. \quad (7.92)$$

This is the most general potential for which we get a conservation law. For $c_0 = 0$ and $\gamma = e^2$ we have the hydrogen atom potential

$$V(r) = -\frac{e^2}{r}, \quad (7.93)$$

so we have

$$\frac{d}{dt} \left(\mathbf{p} \times \mathbf{L} - me^2 \frac{\mathbf{r}}{r} \right) = 0. \quad (7.94)$$

Factoring a constant we obtain the unit-free conserved **Runge-Lenz** vector \mathbf{R} :

$$\mathbf{R} \equiv \frac{1}{me^2} \mathbf{p} \times \mathbf{L} - \frac{\mathbf{r}}{r}, \quad \frac{d\mathbf{R}}{dt} = 0.$$

(7.95)

The conservation of the Runge-Lenz vector is a property of inverse squared central forces. The second vector in \mathbf{R} is simply minus the unit radial vector.

To understand the Runge-Lenz vector, we first examine its value for a circular orbit, as shown in figure 9. The vector \mathbf{L} is out of the page and $\mathbf{p} \times \mathbf{L}$ points radially outward. The vector \mathbf{R} is thus a competition between the outward radial first term and the inner radial second term. If these two terms would not cancel, the result would be a radial vector (outwards or inwards) but in any case, not conserved, as it rotates with the particle. Happily, the two terms cancel. Indeed for a circular orbit

$$m \frac{v^2}{r} = \frac{e^2}{r^2} \rightarrow \frac{m^2 v^2 r}{me^2} = 1 \rightarrow \frac{(mv)(mvr)}{me^2} = 1 \rightarrow \frac{pL}{me^2} = 1, \quad (7.96)$$

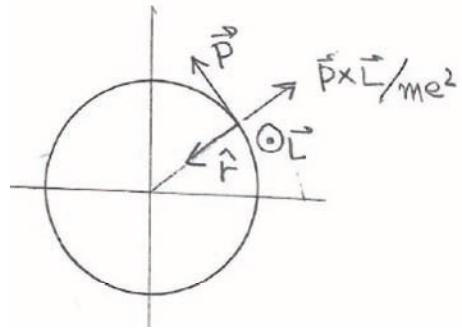


Figure 9: The Runge-Lenz vector vanishes for a circular orbit.

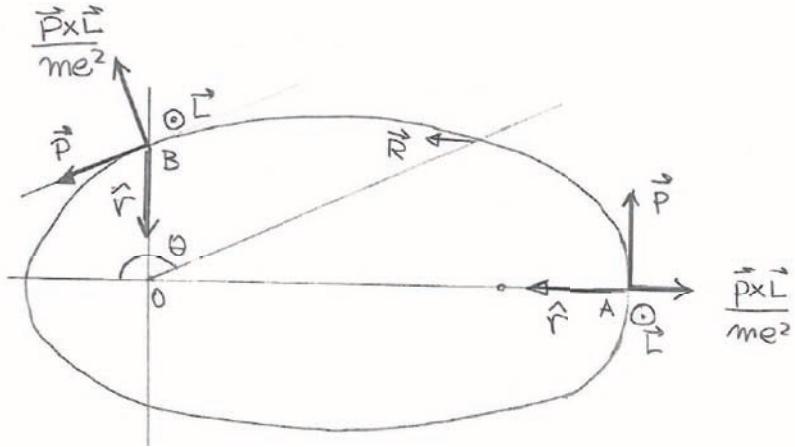


Figure 10: In an elliptic orbit the Runge-Lenz vector is a vector along the major axis of the ellipse and points in the direction from the aphelion to the focus.

which is the statement that the first vector in \mathbf{R} , for a circular orbit, is of unit length and being outward directed cancels with the second term. The Runge-Lenz vector indeed vanishes for a circular orbit.

We now argue that for an elliptical orbit the Runge-Lenz vector is not zero. Consider figure 10. At the aphelion (point furthest away from the focal center), denoted as point A we have the first term in \mathbf{R} point outwards and the second term point inwards. Thus, if \mathbf{R} does not vanish it must be a vector along the line joining the focus and the aphelion, a horizontal vector on the figure. Now consider point B right above the focal center of the orbit. At this point \mathbf{p} is no longer perpendicular to the radial vector and therefore $\mathbf{p} \times \mathbf{L}$ is no longer radial. As you can see, it points slightly to the left. It follows that \mathbf{R} points to the left side of the figure. \mathbf{R} is a vector along the major axis of the ellipse and points in the direction from the aphelion to the focus.

To see more quantitatively the role of \mathbf{R} we dot its definition with the radial vector \mathbf{r} :

$$\mathbf{r} \cdot \mathbf{R} = \frac{1}{me^2} \mathbf{r} \cdot (\mathbf{p} \times \mathbf{L}) - r \quad (7.97)$$

referring to the figure, with the angle θ as defined there and $R \equiv |\mathbf{R}|$, we get

$$rR \cos \theta = \frac{1}{me^2} \mathbf{L} \cdot (\mathbf{r} \times \mathbf{p}) - r = \frac{1}{me^2} L^2 - r. \quad (7.98)$$

Collecting terms proportional to r :

$$r(1 + R \cos \theta) = \frac{L^2}{me^2} \rightarrow \boxed{\frac{1}{r} = \frac{me^2}{L^2}(1 + R \cos \theta)}, \quad (7.99)$$

We identify the magnitude R of the Runge-Lenz vector with the eccentricity of the orbit! Indeed if $R = 0$ the orbit is circular: r does not depend on θ .

This whole analysis has been classical. Quantum mechanically we will need to change some things a bit. The definition of \mathbf{R} only has to be changed to guarantee that \mathbf{R} is a hermitian (vector) operator. As you will verify the hermitization gives

$$\mathbf{R} \equiv \frac{1}{2me^2} (\mathbf{p} \times \mathbf{L} - \mathbf{L} \times \mathbf{p}) - \frac{\mathbf{r}}{r} \quad (7.100)$$

The quantum mechanical conservation of \mathbf{R} is the statement that it commutes with the hydrogen Hamiltonian

$$[\mathbf{R}, H] = 0. \quad (7.101)$$

You will verify this; it is the analog of our classical calculation that showed that the time-derivative of \mathbf{R} is zero. Moreover, the length-squared of the vector is also of interest. You will show that

$$\mathbf{R}^2 = 1 + \frac{2}{me^4} H(\mathbf{L}^2 + \hbar^2). \quad (7.102)$$

MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

PARTICLE NATURE OF LIGHT AND WAVE NATURE OF MATTER

B. Zwiebach
February 16, 2016

Contents

1 Photoelectric Effect	1
2 Compton Scattering	3
3 Matter Waves	6

1 Photoelectric Effect

The photoelectric effect was first observed by Heinrich Hertz in 1887. When polished metal plates are irradiated, they may emit electrons, then called “photo-electrons”. The emitted electrons thus produce a *photoelectric current*. The key observations were:

- There is a threshold frequency ν_0 . Only for frequencies $\nu > \nu_0$ is there a photoelectric current. The frequency ν_0 depends on the metal and the configuration of the atoms at the surface. It is also affected by inhomogeneities.
- The magnitude of the photoelectric current is proportional to the intensity of the light source.
- Energy of the photoelectrons is *independent* of the intensity of the light source.

A natural explanation for the features in this effect didn't come until 1905, when Einstein explained the above features by postulating that the energy in light is carried by discrete quanta (later called photons) with energy $h\nu$. Here h is Planck's constant, the constant used by Planck to produce a fit for the blackbody energy as a function of frequency.

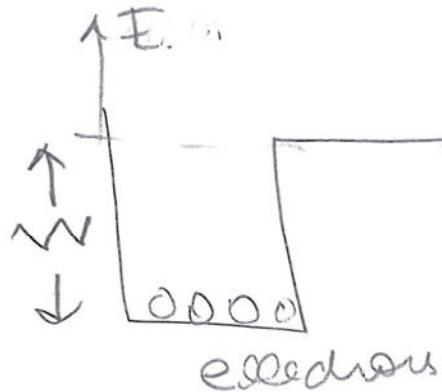


Figure 1: Electrons in a metal are bound. If the photon energy is greater than the work function W an electron may be ejected.

A given material has a characteristic energy W , called the *work function*, which is the minimum energy required to eject an electron. This is not easily calculated because it is the result of an

interaction of many electrons with the background of atoms. It is easily measured, however. When the surface of the material is irradiated, electrons in the material absorb the energy of the incoming photons. If the energy imparted on an electron by the absorption of a single photon is greater than the work function W , then the electron is ejected with kinetic energy E_{e^-} equal to the difference of the photon energy and the work function:

$$E_{e^-} = \frac{1}{2}mv^2 = h\nu - W = E_\gamma - W. \quad (1.1)$$

This equation, written by Einstein explains the experimental features noted above, once we assume that the quanta act on individual electrons to eject them. The threshold frequency is defined by

$$h\nu_0 = W, \quad (1.2)$$

as it leads to a photoelectron with zero energy. For $\nu > \nu_0$ the electrons will be ejected. Increasing the intensity of the light source increases the rate that photons arrive, which will increase the magnitude of the current, but will not change the energy of the photoelectrons because it does not change the energy of each incoming quanta.

Equation (1.2) allowed Einstein to make a prediction: The kinetic energy of the photo-electrons increases linearly with the frequency of light. Einstein's prediction was confirmed experimentally by Millikan (1915) who measured carefully the photoelectron energies and confirmed their linear dependence on the energy. Millikan's careful work allowed him to determine the value of Planck's constant \hbar to better than 1% accuracy! Still, skepticism remained and physicists were not yet convinced about the particle nature of these light quanta.

Example: Consider UV light with wavelength $\lambda = 290\text{nm}$ incident on a metal with work function $W = 4.05\text{eV}$. What is the energy of the photo-electron and what is its speed?

Solution: It is useful to solve these problems without having to look up constants. For this try recalling this useful relation

$$\hbar c = 197.33 \text{ MeV.fm}, \quad \hbar \equiv \frac{h}{2\pi}, \quad (1.3)$$

where $\text{MeV} = 10^6\text{eV}$ and $\text{fm} = 10^{-15}\text{m}$. Let us use this to compute the photon energy. In this case,

$$E_\gamma = h\nu = 2\pi\hbar\frac{c}{\lambda} = \frac{2\pi \cdot 197.33 \text{ MeV.fm}}{290 \times 10^{-9}\text{m}} = \frac{2\pi \cdot 197.33}{290} \text{ eV} \approx 4.28 \text{ eV}, \quad (1.4)$$

and thus

$$E_{e^-} = E_\gamma - W = 0.23 \text{ eV}. \quad (1.5)$$

To compute the energy we set

$$0.23 \text{ eV} = \frac{1}{2}m_e v^2 = \frac{1}{2}(m_e c^2) \left(\frac{v}{c}\right)^2 \quad (1.6)$$

Recalling that $m_e c^2 \simeq 511,000\text{eV}$ one finds

$$\frac{0.46}{511000} = \left(\frac{v}{c}\right)^2 \rightarrow \frac{v}{c} = 0.0009488. \quad (1.7)$$

With and $c = 300,000 \text{ Km/s}$ we finally get $v \simeq 284.4 \text{ Km/s}$.

This is a good point to consider units, in particular the units of h . We can ask: Is there a physical quantity that has the units of h . The answer is yes, as we will see now. From the equation $E = h\nu$, we have

$$[h] = \left[\frac{E}{\nu}\right] = \frac{ML^2/T^2}{1/T} = L \cdot M \frac{L}{T}, \quad (1.8)$$

where $[\cdot]$ gives the units of a quantity, and M, L, T are units of mass, length, and time, respectively. We have written the right-most expression as a product of units of length and momentum. Therefore

$$[h] = [\mathbf{r} \times \mathbf{p}] = [\mathbf{L}]. \quad (1.9)$$

We see that h has units of angular momentum! Indeed for a spin one-half particle, the magnitude of the spin angular momentum is $\frac{1}{2}\hbar$.

With $[h] = [r][p]$ we also see that one has a canonical way to associate a length to any particle of a given mass m . Indeed, using the speed of light, we can construct the momentum $p = mc$, and then the length ℓ is obtained from the ratio h/p . This actually is the **Compton wavelength** λ_C of a particle:

$$\lambda_C = \frac{h}{mc} \quad (1.10)$$

then has units of length; this is called the *Compton wavelength* of a particle of mass m . Note that this length is independent of the velocity of the particle. The de Broglie wavelength of the particle uses the true momentum of the particle, not mc ! Thus, Compton and de Broglie wavelengths should not be confused!

It is possible to get some physical intuition for the Compton wavelength λ_C of a particle. We claim that λ_C is the wavelength of a photon whose energy is equal to the rest energy of the particle. Indeed we would have

$$mc^2 = h\nu = h\frac{c}{\lambda} \rightarrow \lambda = \frac{h}{mc}, \quad (1.11)$$

confirming the claim. Suppose you are trying to localize a point particle of mass m . If you use light, the possible accuracy in the position of the particle is roughly the wavelength of the light. Once we use light with $\lambda < \lambda_C$ the photons carry more energy than the rest energy of the particle. It is possible then that the energy of the photons go into creating more particles of mass m , making it difficult, if not impossible to localize the particle. The Compton wavelength is the length scale at which we need *relativistic quantum field theory* to take into account the possible processes of particle creation and annihilation.

Let us calculate the Compton wavelength of the electron:

$$\lambda_C(e) = \frac{h}{m_e c} = \frac{2\pi\hbar c}{m_e c^2} = \frac{2\pi \cdot 197.33 \text{ MeV.fm}}{0.511 \text{ MeV}} = 2426 \text{ fm} = 2.426 \text{ pm}. \quad (1.12)$$

This length is about 20 times smaller than the Bohr radius (53 pm.) and about two-thousand times the size of a proton (1 fm.). The Compton wavelength of the electron appears in the formula for the change of photon wavelength in the process called Compton scattering.

2 Compton Scattering

Originally Einstein did not make clear that the light quantum meant a particle of light. In 1916, however, he posited that the quantum would carry momentum as well as energy, making the case for a particle much clearer. In relativity, the energy, momentum, and rest mass of a particle are related by

$$E^2 - p^2 c^2 = m^2 c^4. \quad (2.13)$$

(Compare this with the classical equation $E = p^2/2m$.) Of course, one can also express the energy and momentum of the particle in terms of the velocity:

$$E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad \mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (2.14)$$

You should use these expressions to confirm that (2.13) holds ($|\mathbf{p}| = p$). A particle that moves with the speed of light, like the photon, must have zero rest mass, otherwise its energy and momentum would be infinite due to the vanishing denominators. With the rest mass set to zero, equation (2.13) gives the relation between the photon energy E_γ and the photon momentum p_γ :

$$E_\gamma = p_\gamma c. \quad (2.15)$$

Then, using $\lambda\nu = c$, we reach

$$p_\gamma = \frac{E_\gamma}{c} = \frac{h\nu}{c} = \frac{h}{\lambda}. \quad (2.16)$$

We will see this relation again later when we discuss matter waves.

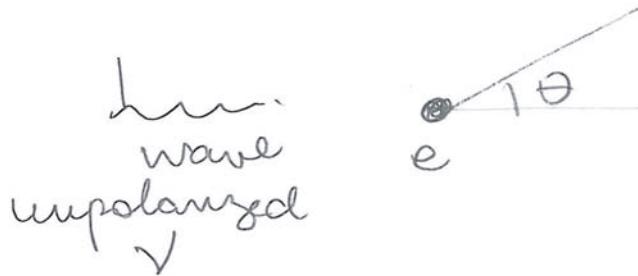


Figure 2: Unpolarized light incident on an electron scatters into an angle θ . Classically, this is described by Thomson scattering. The light does not change frequency during this process.

Compton carried out experiments (1923–1924) scattering X-rays off a carbon target. X-rays correspond to photon energies in the range from 100 eV to 100 KeV. The goal was scattering X-ray photons off free electrons, and with some qualification, the electrons in the atoms behave this way.

The classical counterpart of the Compton experiment is the scattering of electromagnetic waves off free electrons, called *Thompson scattering*. Here an electromagnetic wave is incident on an electron. The electric field of the wave shakes the electron which oscillates with the frequency of the incoming field. The electron oscillation produces a radiated field, of the same frequency as that of the incoming radiation. In classical Thomson scattering the differential scattering cross section is given by

$$\frac{d\sigma}{d\Omega} = \left(\frac{e^2}{mc^2} \right)^2 \frac{1}{2} (1 + \cos^2 \theta), \quad (2.17)$$

where θ is the angle between the incident and scattered wave, with the radiated energy at the same frequency as the incoming light. This is shown in Figure 2. The cross-section has units of length-squared, or area, as it should. It represents the area that would extract from the incoming plane wave the amount of energy that is scattered by the electron. Indeed the quantity $e^2/(mc^2)$ is called the classical electron radius and it is about 2.8 fm! not much bigger than a proton!

If we treat the light as photons, the elementary process going on is a collision between two particles; an incoming photon and a roughly stationary electron. Two facts can be quickly demonstrated:

- The photon cannot be absorbed by the electron. It is inconsistent with energy and momentum conservation (exercise)
- The photon must lose some energy and thus the final photon wavelength λ_f must be larger than the initial photon wavelength λ_i . This is clear in the laboratory frame, where the initially stationary electron must recoil and thus acquire some kinetic energy.

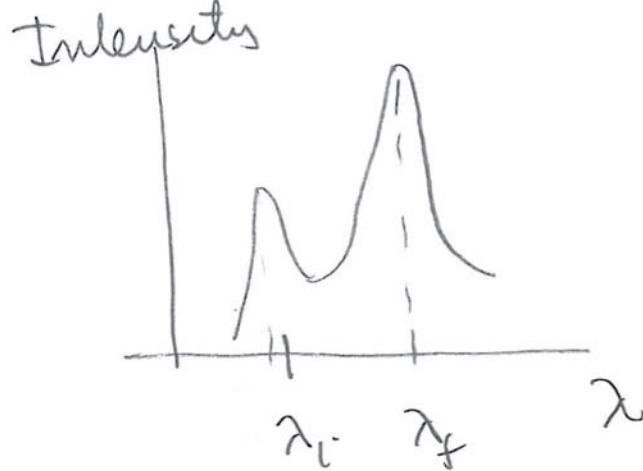


Figure 3: The results of Compton's scattering experiment. The incident photon wavelength is λ_i , and the scattered photon wavelength is $\lambda_f \simeq \lambda_i + \ell_C$, corresponding to $\theta = 90^\circ$.

Indeed, Compton's observations did not agree with the predictions of Thompson scattering: the X-rays changed frequency after scattering. A calculation using energy and momentum conservation shows that the change of wavelength is correlated with the angle between the scattered photon and the original photon:

$$\lambda_f = \lambda_i + \frac{h}{m_e c} (1 - \cos \theta) = \lambda_i + \ell_C (1 - \cos \theta). \quad (2.18)$$

Note that appearance of the Compton wavelength of the electron, the particle the photon scatters off from. The maximum energy loss for the photon occurs at $\theta = \pi$, where

$$\lambda_f(\theta = 180^\circ) = \lambda_i + 2\lambda_C. \quad (2.19)$$

The maximum possible change in wavelength is $2\lambda_C$. For $\theta = \frac{\pi}{2}$ the change of wavelength is exactly ℓ_C

$$\lambda_f(\theta = 90^\circ) = \lambda_i + \lambda_C. \quad (2.20)$$

Compton's experiment used molybdenum X-rays with energy and wavelength

$$E_\gamma \approx 17.5 \text{ keV}, \quad \lambda_i = 0.0709 \text{ nm}, \quad (2.21)$$

incident on a carbon target. Placing the detector at an angle $\theta = 90^\circ$ the plot of the intensity (or number of photons scattered) as a function of wavelength is shown in Figure 2. One finds a peak for $\lambda_f = 0.0731 \text{ nm}$, but also a second peak at the original wavelength $\lambda_i = 0.0709 \text{ nm}$.

The peak at λ_f is the expected one: $\lambda_f - \lambda_i \simeq 2.2 \text{ pm}$, which is about the Compton wavelength of 2.4 pm. Given that the photons have energies of 17 KeV and the bound state energies of carbon

are about 300 eV, the expected peak represents instances where the atom is ionized by the collision and it is a fine approximation to consider the ejected electrons. The peak at λ_i represents a process in which an electron receives some momentum from the photon but still remains bound. This is not very unlikely: the typical momentum of a bound electron is actually comparable to the momentum of the photon. In this case the photon scatters at 90° and the recoil momentum is carried by the whole atom. The relevant Compton wavelength is therefore that of the atom. Since the mass of the carbon atom is several thousands of times larger than the mass of the electron, the Compton wavelength of the atom is much smaller than the electron Compton wavelength and there should be no detectable change in the wavelength of the photon.¹

3 Matter Waves

As we have seen, light behaves as both a particle and a wave. This kind of behavior is usually said to be a **duality**: the complete reality of the object is captured using *both* the wave and particle features of the object. The photon is a particle of energy E_γ , but has frequency ν which is a wave attribute, with $E = h\nu$. It is a particle with momentum p_γ but it also has a wavelength λ , a wave attribute, given by (2.16)

$$\lambda = \frac{h}{p_\gamma}. \quad (3.22)$$

In 1924, Louis de Broglie proposed that the wave/particle duality of the photon was universal, and thus valid for matter particles too. In this way he conjectured the *wave nature of matter*. Inspired by (3.22) de Broglie postulated that associated to a matter particle with momentum p there is a plane wave of wavelength λ given by

$$\lambda = \frac{h}{p}. \quad (3.23)$$

This is a fully quantum property: if $h \rightarrow 0$, then $\lambda \rightarrow 0$, and the particles have no wave properties. And exciting consequence of this is that matter particles can diffract or interfere! In the famous Davisson-Germer experiment (1927) electrons are strike a metal surface and one finds that at certain angles there are peaks in the intensity of the scattered electrons. The peaks showed the effect of constructive interference from scattering off the lattice of atoms in the metal, demonstrating the wave nature of the electrons. One can also do two-slit interference with electrons, and the experiment can be done shooting one electron at a time. A recent experiment [arXiv:1310.8343] by Eibenberger *et.al* reports interference using molecules with 810 atoms and mass exceeding 10 000 amu (that's 20 million times the mass of the electron!).

The de Broglie wavelength can be calculated to estimate if quantum effects are important. Consider for this purpose a particle of mass m and momentum p incident upon an object of size x , as illustrated in Figure 3. Let $\lambda = h/p$ denote the de Broglie wavelength of the particle. The wave nature of the particle is not important if λ is much smaller than x . Thus, the “classical approximation,” in which wave effects are negligible, requires

$$\text{Wave effects negligible: } \frac{\lambda}{x} \ll 1. \quad (3.24)$$

Using $\lambda = h/p$, this yields

$$\text{Wave effects negligible: } x p \gg h, \quad (3.25)$$

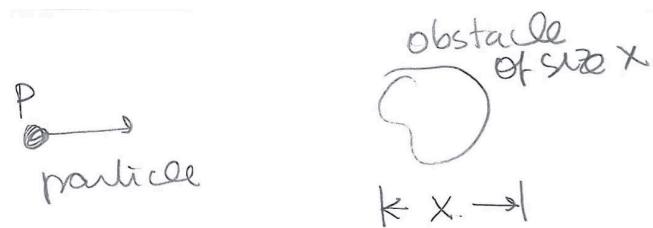


Figure 4: A particle of momentum p incident on an obstacle of size x .

a relation in which both sides have units of angular momentum.

Classical behavior is a subtle limit of quantum mechanics: a classical electromagnetic field requires a large number of photons. Any state with an exact, fixed number of photons, even if large, is not classical, however. Classical electromagnetic states are so-called coherent states, in which the number of photons fluctuates.

Andrew Turner transcribed Zwiebach's handwritten notes to create the first LaTeX version of this document.

¹Thanks to V. Vuletic for a clarification of this point.

MIT OpenCourseWare

<https://ocw.mit.edu>

8.04 Quantum Physics I

Spring 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.

MULTIPARTICLE STATES AND TENSOR PRODUCTS

B. Zwiebach

December 14, 2013

Contents

1	Introduction to the Tensor Product	1
2	Entangled States	8
3	Bell basis states	9
4	Quantum Teleportation	13
5	EPR and Bell Inequalities	16

1 Introduction to the Tensor Product

In this section, we develop the tools needed to describe a system that contains more than one particle. Most of the required ideas appear when we consider systems with two particles. We will assume the particles are distinguishable; for indistinguishable particles quantum mechanics imposes some additional constraints on the allowed set of states. We will study those constraints later in the course (or in 8.06!) The tools we are about to develop will be needed to understand addition of angular momenta. In that problem one is adding the angular momenta of the two or more particles in the system.

Consider then two particles. Below is a description of the quantum mechanics and family of operators associated with each particle:

- Particle 1: its quantum mechanics is described by a complex vector space V . It has associated operators T_1, T_2, \dots
- Particle 2: its quantum mechanics is described by a complex vector space W . It has associated operators S_1, S_2, \dots

This list of operators for each particle may include some or many of the operators you are already familiar with: position, momentum, spin, Hamiltonians, projectors, etc.

Once we have two particles, the two of them together form our system. We are after the description of quantum states of this two-particle system. On first thought, we may think that any state of this system should be described by giving the state $v \in V$ of the first particle and the state $w \in W$ of the second particle. This information could be represented by the ordered list (v, w) where the first item

is the state of the first particle and the second item the state of the second particle. This *is* a state of the two-particle system, but it is far from being the general state of the two-particle system. It misses remarkable new possibilities, as we shall soon see.

We thus introduce a new notation. Instead of representing the state of the two-particle system with particle one in v and particle two in w as (v, w) , we will represent it as $v \otimes w$. This element $v \otimes w$ will be viewed as a vector in a new vector space $V \otimes W$ that will carry the description of the quantum states of the system of two particles. This \otimes operation is called the “tensor product.” In this case we have two vector spaces over \mathbb{C} and the tensor product $V \otimes W$ is a new complex vector space:

$$v \otimes w \in V \otimes W \quad \text{when } v \in V, w \in W. \quad (1.1)$$

In $v \otimes w$ there is no multiplication to be carried out, we are just placing one vector to the left of \otimes and another to the right of \otimes .

We have only described some elements of $V \otimes W$, not quite given its definition yet.¹ We now explain two physically motivated rules that define the tensor product completely.

1. If the vector representing the state of the first particle is scaled by a complex number this is equivalent to scaling the state of the two particles. The same for the second particle. So we declare

$$(av) \otimes w = v \otimes (aw) = a(v \otimes w), \quad a \in \mathbb{C}. \quad (1.2)$$

2. If the state of the first particle is a superposition of two states, the state of the two-particle system is also a superposition. We thus demand distributive properties for the tensor product:

$$\begin{aligned} (v_1 + v_2) \otimes w &= v_1 \otimes w + v_2 \otimes w, \\ v \otimes (w_1 + w_2) &= v \otimes w_1 + v \otimes w_2. \end{aligned} \quad (1.3)$$

The tensor product $V \otimes W$ is thus defined to be the vector space whose elements are (complex) linear combinations of elements of the form $v \otimes w$, with $v \in V, w \in W$, with the above rules for manipulation. The tensor product $V \otimes W$ is the complex vector space of states of the two-particle system!

Comments

1. The vector $0 \in V \otimes W$ is equal to $0 \otimes w$ or $v \otimes 0$. Indeed, by the first property above, with $a = 0$, we have $av = 0$ (rhs a vector) and $0 \otimes w = 0(0 \otimes w) = 0$
2. Let $v_1, v_2 \in V$ and $w_1, w_2 \in W$. A vector in $V \otimes W$ constructed by superposition is

$$\alpha_1(v_1 \otimes w_1) + \alpha_2(v_2 \otimes w_2) \in V \otimes W \quad (1.4)$$

¹If we just left it like this, we would have defined the direct product of vector spaces.

This shows clearly that a general state of the two-particle system cannot be described by stating the state of the first particle and the state of the second particle. The above superpositions give rise to entangled states. An entangled state of the two particles is one that, roughly, cannot be disentangled into separate states of each of the particles. We will make this precise soon.

If (e_1, \dots, e_n) is a basis of V and (f_1, \dots, f_m) is a basis of W , then the set of elements $e_i \otimes f_j$ where $i = 1, \dots, n$ and $j = 1, \dots, m$ forms a basis for $V \otimes W$. It is simple to see these span the space since for any $v \otimes w$ we have $v = \sum_i v_i e_i$ and $w = \sum_j w_j f_j$ so that

$$v \otimes w = (\sum_i v_i e_i) \otimes (\sum_j w_j f_j) = \sum_{i,j} v_i w_j e_i \otimes f_j. \quad (1.5)$$

Given this, we see that the basis also spans linear superpositions of elements of the form $v \otimes w$, thus general elements of $V \otimes W$. With $n \cdot m$ basis vectors, the dimensionality of $V \otimes W$ is equal to the *product* of the dimensionalities of V and W :

$$\dim(V \otimes W) = \dim(V) \times \dim(W). \quad (1.6)$$

Dimensions are multiplied (not added) in a tensor product.

How do we construct operators that act in the vector space $V \otimes W$? Let T be an operator in V and S be an operator in W . In other words, $T \in \mathcal{L}(V)$ and $S \in \mathcal{L}(W)$. We can then construct an operator $T \otimes S$

$$T \otimes S \in \mathcal{L}(V \otimes W) \quad (1.7)$$

defined to act as follows:

$$T \otimes S (v \otimes w) \equiv T v \otimes S w. \quad (1.8)$$

This is the only ‘natural’ option: we let T act on the vector it knows how to act, and S act on the vector it knows how to act.

Suppose that we want the operator $T \in \mathcal{L}(V)$ that acts on the first particle to act on the tensor product $V \otimes W$, even though we have not supplied an operator S to act on the W part. For this we upgrade the operator from one that acts on a single vector space to one, given by $T \otimes 1$, that acts on the tensor product:

$$T \in \mathcal{L}(V) \rightarrow T \otimes 1 \in \mathcal{L}(V \otimes W), \quad T \otimes 1 (v \otimes w) \equiv T v \otimes w. \quad (1.9)$$

Similarly, an operator S belonging to $\mathcal{L}(W)$ is upgraded to $1 \otimes S$ to act on the tensor product. A basic result is that upgraded operators of the first particle **commute** with upgraded operators of the second particle. Indeed,

$$\begin{aligned} (T \otimes 1) \cdot (1 \otimes S) (v \otimes w) &= (T \otimes 1)(v \otimes S w) = T v \otimes S w \\ (1 \otimes S) \cdot (T \otimes 1) (v \otimes w) &= (1 \otimes S)(T v \otimes w) = T v \otimes S w. \end{aligned} \quad (1.10)$$

and therefore

$$[T \otimes \mathbf{1}, \mathbf{1} \otimes S] = 0. \quad (1.11)$$

Given a system of two particles we can construct a simple total Hamiltonian H_T (describing no interactions) by upgrading each of the Hamiltonians H_1 and H_2 and adding them:

$$H_T \equiv H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2 \quad (1.12)$$

Exercise. Convince yourself that

$$\exp\left(-\frac{iH_T t}{\hbar}\right) = \exp\left(-\frac{iH_1 t}{\hbar}\right) \otimes \exp\left(-\frac{iH_2 t}{\hbar}\right) \quad (1.13)$$

We turn now to a famous example at the basis of adding angular momenta.

Example 1: We have two spin-1/2 particles, and describe the first's state space V_1 with basis states $|+\rangle_1$ and $|-\rangle_1$ and the second's state space V_2 with basis states $|+\rangle_2$ and $|-\rangle_2$. The tensor product $V_1 \otimes V_2$ has four basis vectors:

$$|+\rangle_1 \otimes |+\rangle_2; \quad |+\rangle_1 \otimes |-\rangle_2; \quad |-\rangle_1 \otimes |+\rangle_2; \quad |-\rangle_1 \otimes |-\rangle_2 \quad (1.14)$$

If we follow the convention that the first ket corresponds to particle one and the second ket corresponds to particle two, the notation is simpler. The most general state of the two-particle system is a linear superposition of the four basis states:

$$|\Psi\rangle = \alpha_1|+\rangle_1 \otimes |+\rangle_2 + \alpha_2|+\rangle_1 \otimes |-\rangle_2 + \alpha_3|-\rangle_1 \otimes |+\rangle_2 + \alpha_4|-\rangle_1 \otimes |-\rangle_2. \quad (1.15)$$

Example 2: We now want to act on this state with the *total z-component* of angular momentum. Naively, this would be the sum of the z -components of each individual particle. However, we know better at this point - summing the two angular momenta really means constructing a new operator in the tensor product vector space:

$$S_z^T = S_z^{(1)} \otimes \mathbf{1} + \mathbf{1} \otimes S_z^{(2)}. \quad (1.16)$$

Performing the calculation in two parts,

$$\begin{aligned} (S_z^{(1)} \otimes \mathbf{1})|\Psi\rangle &= \alpha_1 S_z |+\rangle_1 \otimes |+\rangle_2 + \alpha_2 S_z |+\rangle_1 \otimes |-\rangle_2 + \alpha_3 S_z |-\rangle_1 \otimes |+\rangle_2 + \alpha_4 S_z |-\rangle_1 \otimes |-\rangle_2 \\ &= \frac{\hbar}{2} \left(\alpha_1 |+\rangle_1 \otimes |+\rangle_2 + \alpha_2 |+\rangle_1 \otimes |-\rangle_2 - \alpha_3 |-\rangle_1 \otimes |+\rangle_2 - \alpha_4 |-\rangle_1 \otimes |-\rangle_2 \right) \\ (1 \otimes S_z^{(2)})|\Psi\rangle &= \alpha_1 |+\rangle_1 \otimes S_z |+\rangle_2 + \alpha_2 |+\rangle_1 \otimes S_z |-\rangle_2 + \alpha_3 |-\rangle_1 \otimes S_z |+\rangle_2 + \alpha_4 |-\rangle_1 \otimes S_z |-\rangle_2 \\ &= \frac{\hbar}{2} \left(\alpha_1 |+\rangle_1 \otimes |+\rangle_2 - \alpha_2 |+\rangle_1 \otimes |-\rangle_2 + \alpha_3 |-\rangle_1 \otimes |+\rangle_2 - \alpha_4 |-\rangle_1 \otimes |-\rangle_2 \right) \end{aligned} \quad (1.17)$$

Adding these together, we have:

$$S_z^T |\Psi\rangle = \hbar (\alpha_1 |+\rangle_1 \otimes |+\rangle_2 - \alpha_4 |-\rangle_1 \otimes |-\rangle_2) \quad (1.18)$$

One can derive this result quickly by noting that since $S_z^{(1)}$ is diagonal in the first basis and $S_z^{(2)}$ is diagonal in the second basis, the total S_z is diagonal in the tensor space basis and its eigenvalue acting on a tensor state is the sum of the S_z eigenvalues for particle one and particle two. Thus,

$$\begin{aligned} S_z^T |+\rangle \otimes |+\rangle &= \left(\frac{\hbar}{2} + \frac{\hbar}{2}\right) |+\rangle \otimes |+\rangle = \hbar |+\rangle \otimes |+\rangle \\ S_z^T |+\rangle \otimes |-\rangle &= \left(\frac{\hbar}{2} - \frac{\hbar}{2}\right) |+\rangle \otimes |+\rangle = 0 \\ S_z^T |-\rangle \otimes |+\rangle &= \left(-\frac{\hbar}{2} + \frac{\hbar}{2}\right) |+\rangle \otimes |+\rangle = 0 \\ S_z^T |-\rangle \otimes |-\rangle &= \left(-\frac{\hbar}{2} - \frac{\hbar}{2}\right) |-\rangle \otimes |-\rangle = -\hbar |-\rangle \otimes |-\rangle \end{aligned} \quad (1.19)$$

The result in (1.18) follows quickly from the four relations above. Suppose we are only interested in states that have zero S_z^T . This requires

$$\alpha_1 = \alpha_4 = 0 \rightarrow |\Psi\rangle = \alpha_2 |+\rangle \otimes |-\rangle + \alpha_3 |-\rangle \otimes |+\rangle \quad (1.20)$$

Example 3: Calculate the total x -component S_x^T of spin angular momentum on the above states with zero S_z^T . Recalling that

$$S_x |+\rangle = \frac{\hbar}{2} |-\rangle, \quad S_x |-\rangle = \frac{\hbar}{2} |+\rangle \quad (1.21)$$

and writing

$$S_x^T = S_x \otimes 1 + 1 \otimes S_x \quad (1.22)$$

the calculation proceeds as follows:

$$\begin{aligned} S_x^T |+\rangle \otimes |-\rangle &= S_x |+\rangle \otimes |-\rangle + |+\rangle \otimes S_x |-\rangle = \frac{\hbar}{2} (|-\rangle \otimes |-\rangle + |+\rangle \otimes |+\rangle) \\ S_x^T |-\rangle \otimes |+\rangle &= S_x |-\rangle \otimes |+\rangle + |-\rangle \otimes S_x |+\rangle = \frac{\hbar}{2} (|+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle) \end{aligned} \quad (1.23)$$

Therefore

$$\begin{aligned} S_x^T |\Psi\rangle &= \alpha_2 \frac{\hbar}{2} (|-\rangle \otimes |-\rangle + |+\rangle \otimes |+\rangle) + \alpha_3 \frac{\hbar}{2} (|+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle) \\ &= \frac{\hbar}{2} (\alpha_2 + \alpha_3) (|+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle) \end{aligned} \quad (1.24)$$

If we demand that S_x^T also be zero on the state we now find $\alpha_2 = -\alpha_3$. Thus, the following state has zero S_x^T, S_z^T :

$$|\Psi\rangle = \alpha (|+\rangle \otimes |-\rangle - |-\rangle \otimes |+\rangle). \quad (1.25)$$

Exercise: Verify that $S_y^T |\Psi\rangle = 0$. Thus we say that the state has total spin angular momentum zero.

We now consider the definition of an **inner product** in $V \otimes W$. To do this we simply give state how the most general inner product is computed using a basis $\{e_i \otimes f_j\}$ for the tensor product, with $\{e_i\}$ and $\{f_i\}$ *orthonormal* bases for V and W . We begin by declaring that

$$\langle e_i \otimes f_j, e_p \otimes f_q \rangle \equiv \delta_{ip} \delta_{jq}. \quad (1.26)$$

This makes the basis $\{e_i \otimes f_j\}$ orthonormal. In addition, we must declare that with vectors $X, Y, Z \in V \otimes W$ and a complex constant a the following axioms hold:

$$\begin{aligned} \langle X + Y, Z \rangle &= \langle X, Z \rangle + \langle Y, Z \rangle, \\ \langle X, Y + Z \rangle &= \langle X, Y \rangle + \langle X, Z \rangle, \\ \langle X, aY \rangle &= a \langle X, Y \rangle \\ \langle aX, Y \rangle &= a^* \langle X, Y \rangle \end{aligned} \quad (1.27)$$

This is a complete definition of the inner product in the tensor space: we can compute the inner product of any two vectors in $V \otimes W$ using the chosen basis and the above distributive rules. Indeed, using these properties we can show that

$$\langle v \otimes w, \tilde{v} \otimes \tilde{w} \rangle = \langle v, \tilde{v} \rangle \langle w, \tilde{w} \rangle, \quad (1.28)$$

where the inner products on the right-hand side are those in V and in W , making it clear that the inner product in $V \otimes W$ arises from the inner products in V and W . To prove this relation we begin by writing

$$\begin{aligned} v &= \sum_i v_i e_i, & w &= \sum_j w_j f_j, \\ \tilde{v} &= \sum_p \tilde{v}_p e_p, & \tilde{w} &= \sum_q \tilde{w}_q f_q. \end{aligned} \quad (1.29)$$

Since the basis vectors in V and W are orthonormal we find that

$$\langle v, \tilde{v} \rangle = \sum_i v_i^* \tilde{v}_i, \quad \langle w, \tilde{w} \rangle = \sum_j w_j^* \tilde{w}_j. \quad (1.30)$$

Now evaluating the left-hand side of (1.28)

$$\begin{aligned} \langle v \otimes w, \tilde{v} \otimes \tilde{w} \rangle &= \left\langle \sum_i v_i e_i \otimes \sum_j w_j f_j, \sum_p \tilde{v}_p e_p \otimes \sum_q \tilde{w}_q f_q \right\rangle \\ &= \sum_{i,j,p,q} \left\langle v_i w_j e_i \otimes f_j, \tilde{v}_p \tilde{w}_q e_p \otimes f_q \right\rangle \\ &= \sum_{i,j,p,q} v_i^* w_j^* \tilde{v}_p \tilde{w}_q \langle e_i \otimes f_j, e_p \otimes f_q \rangle \\ &= \sum_{i,j,p,q} v_i^* w_j^* \tilde{v}_p \tilde{w}_q \delta_{ip} \delta_{jq} = \sum_i v_i^* \tilde{v}_i \sum_j w_j^* \tilde{w}_j \\ &= \langle v, \tilde{v} \rangle \langle w, \tilde{w} \rangle. \end{aligned} \quad (1.31)$$

The verification that the inner-product on $V \otimes W$ satisfies the remaining axioms is left as a good practice for you. Assume below that $X, Y \in V \otimes W$. For both exercises above simply write the most general vector, as $X = \sum_{ij} x_{ij} e_i \otimes f_j$ and proceed.

Exercise: Show that $\langle X, X \rangle \geq 0$, and $\langle X, X \rangle = 0$ if and only if $X = 0$.

Exercise: Show that $\langle X, Y \rangle = \langle Y, X \rangle^*$.

Many times it is convenient to use bra-ket notation for inner products in the tensor product. We write

$$\begin{aligned} |v \otimes w\rangle &= |v\rangle_1 \otimes |w\rangle_2 \\ \langle v \otimes w| &= {}_1\langle v|_1 \otimes {}_2\langle w|. \end{aligned} \quad (1.32)$$

Notice that both on bras and kets we write the state of particle one to the left of the state of particle two. We then write (1.28) as

$$\langle v \otimes w | \tilde{v} \otimes \tilde{w} \rangle = ({}_1\langle v| \otimes {}_2\langle w|) (|\tilde{v}\rangle_1 \otimes |\tilde{w}\rangle_2) = \langle v|\tilde{v}\rangle \langle w|\tilde{w}\rangle. \quad (1.33)$$

Back to our example with spin states, our four basis vectors $|+\rangle_1 \otimes |+\rangle_2$, $|+\rangle_1 \otimes |-\rangle_2$, $|-\rangle_1 \otimes |+\rangle_2$, and $|-\rangle_1 \otimes |-\rangle_2$ are orthonormal. We had the un-normalized state in (1.25) given by

$$|\Psi\rangle = \alpha \left(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2 \right). \quad (1.34)$$

The associated bra is then

$$\langle \Psi| = \alpha^* \left({}_1\langle +| \otimes {}_2\langle -| - {}_1\langle -| \otimes {}_2\langle +| \right). \quad (1.35)$$

We then have

$$\begin{aligned} \langle \Psi | \Psi \rangle &= \alpha \alpha^* \left({}_1\langle +| \otimes {}_2\langle -| - {}_1\langle -| \otimes {}_2\langle +| \right) \left(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2 \right) \\ &= \alpha \alpha^* \left({}_1\langle +| \otimes {}_2\langle -| |+\rangle_1 \otimes |-\rangle_2 + {}_1\langle -| \otimes {}_2\langle +| |-\rangle_1 \otimes |+\rangle_2 \right) \end{aligned} \quad (1.36)$$

since only terms where the spin states are the same for the first particle and for the second particle survive. We thus have, for normalization,

$$\langle \Psi | \Psi \rangle = |\alpha|^2 (1 + 1) = 2|\alpha|^2 = 1, \quad \rightarrow \quad \alpha = \frac{1}{\sqrt{2}}. \quad (1.37)$$

The normalized state with zero total angular momentum is then

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2 \right). \quad (1.38)$$

2 Entangled States

You have learned that $V \otimes W$ includes states $\Psi = \sum_i \alpha_i v_i \otimes w_i$ obtained by linear superposition of simpler states of the form $v_i \otimes w_i$. If handed such a Ψ , you might want to know whether you can write it as a single term $v_* \otimes w_*$ for some $v_* \in V$ and $w_* \in W$. If so, you are able to describe the state of the particles in Ψ independently: particle one is in state v_* and particle two in state w_* . We then say that in the state Ψ the particles are *not entangled*. If no such v_* and w_* exist, we say that in the state $\Psi \in V \otimes W$ the particles are entangled or equivalently, that Ψ is an entangled state of the two particles. Entanglement is a basis-independent property.

It is simplest to illustrate this using two-dimensional complex vector spaces V and W , like the ones we use for spin one-half. Let V have a basis e_1, e_2 and W have a basis f_1, f_2 . Then, the most general state you can write is the following:

$$\Psi_A = a_{11} e_1 \otimes f_1 + a_{12} e_1 \otimes f_2 + a_{21} e_2 \otimes f_1 + a_{22} e_2 \otimes f_2. \quad (2.39)$$

This state is encoded by a matrix A of coefficients

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \quad (2.40)$$

The state is *not entangled* if there exist constants a_1, a_2, b_1, b_2 such that

$$a_{11} e_1 \otimes f_1 + a_{12} e_1 \otimes f_2 + a_{21} e_2 \otimes f_1 + a_{22} e_2 \otimes f_2 = (a_1 e_1 + a_2 e_2) \otimes (b_1 f_1 + b_2 f_2). \quad (2.41)$$

Note that these four unknown constants are not uniquely determined: we can, for example, multiply a_1 and a_2 by some constant $c \neq 0$ and divide b_1 and b_2 by c , to obtain a different solution. Indeed $v \otimes w = (cv) \otimes (w/c)$ for any $c \neq 0$. Using the distributive laws for \otimes to expand the right-hand side of (2.41) and recalling that $e_i \otimes f_j$ are basis vectors in the tensor product, we see that the equality requires the following four relations:

$$\begin{aligned} a_{11} &= a_1 b_1 \\ a_{12} &= a_1 b_2 \\ a_{21} &= a_2 b_1 \\ a_{22} &= a_2 b_2 \end{aligned} \quad (2.42)$$

Combining these four expressions leaves us with a consistency condition:

$$a_{11}a_{22} - a_{12}a_{21} = a_1b_1a_2b_2 - a_1b_2a_2b_1 = 0 \rightarrow \det A = 0. \quad (2.43)$$

In other words, if Ψ_A is not entangled the determinant of the matrix A must be zero. We can in fact show that $\det A = 0$ implies that Ψ_A is not entangled. To do this we simply have to present a solution for the equations above under the condition $\det A = 0$.

Assume first that $a_{11} = 0$. Then $\det A = 0$ implies $a_{12}a_{21} = 0$. If $a_{12} = 0$ then

$$\Psi_A = a_{21}e_2 \otimes f_1 + a_{22}e_2 \otimes f_2 = e_2 \otimes (a_{21}f_1 + a_{22}f_2) \quad (2.44)$$

and the state is indeed not entangled. If $a_{21} = 0$ then

$$\Psi_A = a_{12}e_1 \otimes f_2 + a_{22}e_2 \otimes f_2 = (a_{12}e_1 + a_{22}e_2) \otimes f_2 \quad (2.45)$$

and again, the state is not entangled. Thus, we can solve all equations when $a_{11} = 0$. Now assuming $a_{11} \neq 0$ we can take

$$a_1 = \sqrt{a_{11}}, \quad b_1 = \sqrt{a_{11}}, \quad (2.46)$$

to solve the first equation in (2.42). The second and third equations allow us to solve for b_2 and a_2

$$b_2 = \frac{a_{12}}{\sqrt{a_{11}}}, \quad a_2 = \frac{a_{21}}{\sqrt{a_{11}}} \quad (2.47)$$

The fourth equation is then automatically satisfied as

$$a_2 b_2 = \frac{a_{12}a_{21}}{a_{11}} = \frac{a_{11}a_{22}}{a_{11}} = a_{22} \quad (2.48)$$

using the vanishing determinant condition. We have thus solved the system of equations and we can write

$$\Psi_A = \left(\sqrt{a_{11}}e_1 + \frac{a_{21}}{\sqrt{a_{11}}}e_2 \right) \otimes \left(\sqrt{a_{11}}f_1 + \frac{a_{12}}{\sqrt{a_{11}}}f_2 \right) \quad \text{if } \det A = 0. \quad (2.49)$$

We have thus proved that Ψ_A is entangled if and only if $\det A \neq 0$. For vector spaces of dimensions different than two the conditions for entanglement take a different form. Schrödinger called “entanglement” the essential feature of quantum mechanics.

Example: Consider our state of zero total spin angular momentum:

$$|\Phi\rangle_A \equiv \frac{1}{\sqrt{2}}(|+\rangle_1 \otimes |+\rangle_2 - |-\rangle_1 \otimes |-\rangle_2) \quad (2.50)$$

If we have the basis vectors $|e_1\rangle = |+\rangle_1, |e_2\rangle = |-\rangle_1$ and $|f_1\rangle = |+\rangle_2, |f_2\rangle = |-\rangle_2$ we see that the state is described by the matrix

$$A = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} \end{pmatrix} \quad (2.51)$$

Since the determinant of this matrix is not zero, the state is entangled.

3 Bell basis states

Bell states are a set of entangled basis vectors. Take $V_1 \otimes V_2$, with V_1 and V_2 both the two-dimensional complex vector space of spin-1/2 particles. For brevity of notation we will leave out the 1 and 2

subscripts on the states and the \otimes in between the states; it is always understood that in $V_1 \otimes V_2$ the state of V_1 appears to the left of the state of V_2 . Consider now the state

$$|\Phi_0\rangle \equiv \frac{1}{\sqrt{2}}(|+\rangle|+\rangle + |-\rangle|-\rangle). \quad (3.52)$$

This is clearly an entangled state: its associated matrix is diagonal with equal entries of $1/\sqrt{2}$ and thus non-zero determinant. Moreover this state is unit normalized

$$\langle \Phi_0 | \Phi_0 \rangle = 1. \quad (3.53)$$

We can use this state as the first of our basis vectors for $V_1 \otimes V_2$. Since this tensor product is four-dimensional we need three more entangled basis states. Here they are:

$$|\Phi_i\rangle \equiv (\mathbf{1} \otimes \sigma_i)|\Phi_0\rangle, \quad i = 1, 2, 3. \quad (3.54)$$

We will explicitly see below that these states are entangled, but this property is clear from the definition. If $|\Psi_i\rangle$ is not entangled, it would follow that that $\mathbf{1} \otimes \sigma_i|\Psi_i\rangle$ (i not summed) is not entangled either (do you see why?). But using $\sigma_i^2 = 1$, we see that this last state is in fact $|\Phi_0\rangle$, which is entangled. This contradiction shows that $|\Phi_i\rangle$ must be entangled. It is also manifest from the definition that the $|\Phi_i\rangle$ states are unit normalized.

Let us look at the form of $|\Phi_1\rangle$:

$$\begin{aligned} |\Phi_1\rangle &= (\mathbf{1} \otimes \sigma_1)\frac{1}{\sqrt{2}}(|+\rangle|+\rangle + |-\rangle|-\rangle) = \frac{1}{\sqrt{2}}(|+\rangle\sigma_1|+\rangle + |-\rangle\sigma_1|-\rangle) \\ &= \frac{1}{\sqrt{2}}(|+\rangle|-\rangle + |-\rangle|+\rangle). \end{aligned} \quad (3.55)$$

The state is clearly entangled. By analogous calculations we obtain the full list of Bell states

$$\begin{aligned} |\Phi_0\rangle &= \mathbf{1} \otimes \mathbf{1} |\Phi_0\rangle = \frac{1}{\sqrt{2}}(|+\rangle|+\rangle + |-\rangle|-\rangle) \\ |\Phi_1\rangle &= \mathbf{1} \otimes \sigma_1 |\Phi_0\rangle = \frac{1}{\sqrt{2}}(|+\rangle|-\rangle + |-\rangle|+\rangle) \\ |\Phi_2\rangle &= \mathbf{1} \otimes \sigma_2 |\Phi_0\rangle = \frac{i}{\sqrt{2}}(|+\rangle|-\rangle - |-\rangle|+\rangle) \\ |\Phi_3\rangle &= \mathbf{1} \otimes \sigma_3 |\Phi_0\rangle = \frac{1}{\sqrt{2}}(|+\rangle|+\rangle - |-\rangle|-\rangle). \end{aligned} \quad (3.56)$$

By inspection we can confirm that Φ_0 is orthogonal to the other three: $\langle \Phi_0 | \Phi_i \rangle = 0$. It is not much work either to see that the basis is in fact orthonormal. But a calculation is kind of fun:

$$\begin{aligned} \langle \Phi_i | \Phi_j \rangle &= \langle \Phi_0 | (\mathbf{1} \otimes \sigma_i)(\mathbf{1} \otimes \sigma_j) |\Phi_0\rangle \\ &= \langle \Phi_0 | \mathbf{1} \otimes \sigma_i \sigma_j |\Phi_0\rangle \\ &= \langle \Phi_0 | \mathbf{1} \otimes (\mathbf{1} \delta_{ij} + i \epsilon_{ijk} \sigma_k) |\Phi_0\rangle \\ &= \delta_{ij} \langle \Phi_0 | \mathbf{1} \otimes \mathbf{1} |\Phi_0\rangle + i \epsilon_{ijk} \langle \Phi_0 | \mathbf{1} \otimes \sigma_k |\Phi_0\rangle \\ &= \delta_{ij} \langle \Phi_0 | \Phi_0 \rangle + i \epsilon_{ijk} \langle \Phi_0 | \Phi_k \rangle = \delta_{ij}, \end{aligned} \quad (3.57)$$

as we wanted to show. Indeed, we have an orthonormal basis of entangled states.

We can solve for the old, non-entangled basis states in terms of the Bell states. We quickly find from (3.56)

$$\begin{aligned} |+\rangle|+\rangle &= \frac{1}{\sqrt{2}} (|\Phi_0\rangle + |\Phi_3\rangle) \\ |-\rangle|-\rangle &= \frac{1}{\sqrt{2}} (|\Phi_0\rangle - |\Phi_3\rangle) \\ |+\rangle|-\rangle &= \frac{1}{\sqrt{2}} (|\Phi_1\rangle - i|\Phi_2\rangle) \\ |-\rangle|+\rangle &= \frac{1}{\sqrt{2}} (|\Phi_1\rangle + i|\Phi_2\rangle). \end{aligned} \tag{3.58}$$

Introducing labels A and B for the two spaces in a tensor product $V_A \otimes V_B$ we rewrite the above equations as

$$\begin{aligned} |+\rangle_A|+\rangle_B &= \frac{1}{\sqrt{2}} (|\Phi_0\rangle_{AB} + |\Phi_3\rangle_{AB}) \\ |-\rangle_A|-\rangle_B &= \frac{1}{\sqrt{2}} (|\Phi_0\rangle_{AB} - |\Phi_3\rangle_{AB}) \\ |+\rangle_A|-\rangle_B &= \frac{1}{\sqrt{2}} (|\Phi_1\rangle_{AB} - i|\Phi_2\rangle_{AB}) \\ |-\rangle_A|+\rangle_B &= \frac{1}{\sqrt{2}} (|\Phi_1\rangle_{AB} + i|\Phi_2\rangle_{AB}), \end{aligned} \tag{3.59}$$

where $|\Phi_i\rangle_{AB}$ are the Bell states we defined above with tensor products in which the first state is in V_A and the second state is in V_B .

These basis states form the Bell basis. You could do an experiment to determine the probability of an arbitrary state being along any of the basis states in this orthonormal basis. You can use the experiment to detect which basis state the state is in. The state is, of course, a superposition of basis states, but during measurement will collapse into one of them with some probability. The Stern Gerlach device was an example of a device that allowed you to collapse a state into one basis state or another. This basis is more general, as it is not simply for two-state systems.

We conclude by presenting three facts.

1. Measuring in a basis. Given an orthonormal basis $|e_1\rangle, \dots, |e_n\rangle$ we can measure a state $|\Psi\rangle$ along this basis and obtain that the probability $P(i)$ to be in the state $|i\rangle$ is $|\langle e_i|\Psi\rangle|^2$. After measurement the state will be in one of the states $|e_i\rangle$. This is exactly how it worked for the Stern-Gerlach experiment which, oriented about **z** amount to a measurement in the basis $|+\rangle, |-\rangle$.

As another example, if we have a state with two particles A, B , we may choose the four Bell states as our orthonormal basis for the measurement. If so, after measurement the state will be in one of the Bell states $|\Phi_i\rangle_{AB}$, with probability given by the squared overlap $|\langle \Phi_i|_{AB}|\Psi\rangle|^2$.

2. Partial measurement. Suppose we have a general (entangled) state $\Psi \in V \otimes W$ of two particles. The observer Alice has access to both particles but decides to measure only the first particle along the basis $|e_1\rangle, \dots, |e_n\rangle$ of V . How is this analyzed? As a first step we use that basis to write the state Ψ in the form

$$\Psi = \sum_i |e_i\rangle \otimes |w_i\rangle, \quad (3.60)$$

for some calculable vectors $|w_i\rangle$. As a second step we normalize the states $|w_i\rangle$:

$$\Psi = \sum_i \sqrt{\langle w_i|w_i\rangle} |e_i\rangle \otimes \frac{|w_i\rangle}{\sqrt{\langle w_i|w_i\rangle}}, \quad (3.61)$$

We claim that Alice will find the first particle to be in the state $|i\rangle$ with probability $\langle w_i|w_i\rangle$. After the measurement, the state of the particles will be

$$|e_i\rangle \otimes \frac{|w_i\rangle}{\sqrt{\langle w_i|w_i\rangle}}, \quad \text{for some value of } i. \quad (3.62)$$

(A justification of this answer was given in recitations.) You probably have used this rule before. As an example, suppose we have the entangled state of total spin zero:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2) \quad (3.63)$$

If we measure the first particle along the $|+\rangle_1, |+\rangle_2$ basis we find

$$\begin{aligned} &\text{Probability that the first particle is in } |+\rangle = \frac{1}{2}. \quad \text{State after measurement: } |+\rangle_1 \otimes |-\rangle_2 \\ &\text{Probability that the first particle is in } |-\rangle = \frac{1}{2}. \quad \text{State after measurement: } |-\rangle_1 \otimes |+\rangle_2 \end{aligned} \quad (3.64)$$

It follows that after the measurement of the first particle, a measurement of the second particle will show that its spin is always opposite to the spin of the first particle.

As a more nontrivial example, consider now the state of three particles A, B, C which live in $V_A \otimes V_B \otimes V_C$, which contains states of the type $v \otimes w \otimes u$ with $v \in V_A, w \in V_B, u \in V_C$, and their linear combinations. To analyze what happens if Alice decides to do a Bell measurement of particles A, B , the state Ψ of the system must be written in the form

$$\Psi = |\Phi_0\rangle_{AB} \otimes |u_0\rangle_C + |\Phi_1\rangle_{AB} \otimes |u_1\rangle_C + |\Phi_2\rangle_{AB} \otimes |u_2\rangle_C + |\Phi_3\rangle_{AB} \otimes |u_3\rangle_C \quad (3.65)$$

After the measurement, the state of the particles A, B will be one of the Bell states $|\Phi_\mu\rangle_{AB}$ with $\mu = 0, 1, 2, 3$. We have

$$\begin{aligned} &\text{Probability that } (A, B) \text{ is in } |\Phi_\mu\rangle_{AB} = \langle u_\mu|u_\mu\rangle, \\ &\text{State after measurement is } |\Phi_\mu\rangle_{AB} \otimes \frac{|u_\mu\rangle_C}{\sqrt{\langle u_\mu|u_\mu\rangle}} \text{ for some } \mu \in \{0, 1, 2, 3\}. \end{aligned} \quad (3.66)$$

3. The action of the Pauli matrices on spin states can be realized as time evolution via some Hamiltonian. Note first that the Pauli matrices are unitary because they are Hermitian and square to the identity. Multiplying a state by σ_1 is thus acting with a unitary operator and unitary operators generate allowed time evolution. Thus, there is a Hamiltonian that applied to a system over some length of time will turn any spin state $|\Psi\rangle$ into $\sigma_i|\Psi\rangle$. In practice, this Hamiltonian would correspond to some device with a magnetic field of some determined magnitude and direction that acts for a few picoseconds and evolves spin states in time. We can check, for example, that any Pauli matrix can be written as the exponential of i times a Hermitian matrix (which would be proportional to the Hamiltonian):

$$e^{i\frac{\pi}{2}(-1+\sigma_i)} = e^{-i\frac{\pi}{2}e^{i\frac{\pi}{2}\sigma_i}} = (-i)(i\sigma_i) = \sigma_i \quad (3.67)$$

4 Quantum Teleportation

Classically, teleportation is impossible: there is no classical basis for dematerializing an object and recreating it somewhere else. In 1993, a group of scientists (Bennet, Brassard, Crépeau, Jozsa, Peres, and Wootters) discovered that teleportation *is* possible in quantum mechanics.

Imagine that Alice has a quantum state: the state of a 1/2 particle. The state is:

$$|\Psi\rangle_C = \alpha|+\rangle_C + \beta|-\rangle_C, \quad (4.68)$$

where $\alpha, \beta \in \mathbb{C}$ and the letter C denotes the state space V_C of this C particle to be teleported. Her goal is to teleport this state - called a “quantum bit,” or *qubit* - to Bob, who is far away.

The quantum “no-cloning” principle prevents Alice from simply creating a copy of the state and sending it to Bob. In other words, it is impossible to create a copy of a quantum mechanical state. Measuring the state and telling Bob about the result is no option either: if Alice measures the state with some Stern-Gerlach apparatus, the spin will just point up or point down. What has she learned? Almost nothing. Only with many copies of the state she would be able to learn about the values of α and β . Having just one particle she is unable to measure α and β and send those values to Bob. Of course, it may be that for some reason Alice knows the values of α and β . In some cases she could transmit that information to Bob to recreate the state. But it could also be that α , for example is some transcendental number 0.178573675623..... with no discernible rhyme or reason, and she would need an infinite amount of information to send to Bob this value.

Here is a diagram of how Alice can will teleport the information:

The key tool Alice and Bob use is an entangled states of two particles A and B , in which Alice has access to particle A and Bob has access to particle B . One pair (A, B) of entangled particles will allow Alice to teleport the state C of one particle. To teleport a full person from one place to another, we would have to have an enormous reservoir of entangled pairs, one pair needed to teleport each quantum state of particles in the body of that person. This clearly remains science-fiction.

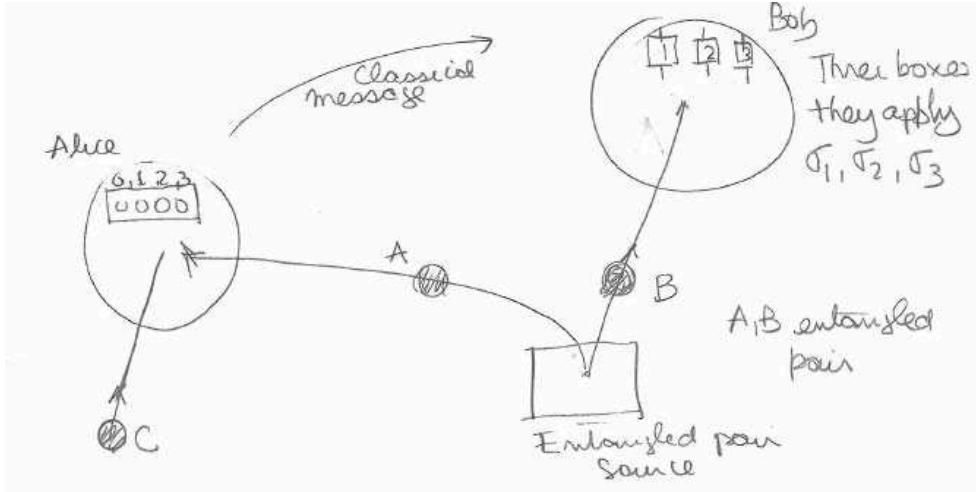


Figure 1: Alice has access to spin state C , to be teleported and to spin state A which is entangled with spin state B . Bob has spin state B . Alice will perform a Bell measurement on states A and C . After she measures, Bob's state will turn into the desired state, up to a simple unitary transformation.

Alice has a console with four lights, labeled with $\mu = 0, 1, 2, 3$. She will do a Bell measurement involving particle A (of the shared entangled pair) and particle C (the one to be teleported). As she does so, one of her four lights will blink: If it is the μ -th light it is because she ended with the Bell state $|\Phi_\mu\rangle_{AC}$. Bob, who is in possession of the particle B , has a console with four boxes which generate unitary transformations (via some Hamiltonians applied for a fixed time). The first box, labeled $\mu = 0$ does nothing to the state. The i -th box (with $i = 1, 2, 3$) applies the operator σ_i . Alice communicates to Bob that the μ -th light blinked. Then Bob submits particle B to the μ -th box and out comes, we claim, the teleported state $|\Psi\rangle_B$ as the state of particle B .

Let us prove this mathematically. Let the entangled shared pair, with A at Alice and B at Bob, be the first Bell basis state:

$$|\Phi_0\rangle_{AB} = \frac{1}{\sqrt{2}}(|+\rangle_A|+\rangle_B + |-\rangle_A|-\rangle_B). \quad (4.69)$$

The total state of our three particles, A, B, C is therefore:

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= |\Phi_0\rangle_{AB} \otimes (\alpha|+\rangle_C + \beta|-\rangle_C) \\ &= \frac{1}{\sqrt{2}}\left(|+\rangle_A|+\rangle_B + |-\rangle_A|-\rangle_B\right) \otimes (\alpha|+\rangle_C + \beta|-\rangle_C). \end{aligned} \quad (4.70)$$

Expanding out and reordering the states to have A followed by C and then by B we have

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= \frac{1}{\sqrt{2}}\left(\underbrace{\alpha|+\rangle_A|+\rangle_C}_{\text{underbrace}}|+\rangle_B + \underbrace{\beta|+\rangle_A|-\rangle_C}_{\text{underbrace}}|+\rangle_B\right. \\ &\quad \left.+ \underbrace{\alpha|-\rangle_A|+\rangle_C}_{\text{underbrace}}|-\rangle_B + \underbrace{\beta|-\rangle_A|-\rangle_C}_{\text{underbrace}}|-\rangle_B\right). \end{aligned} \quad (4.71)$$

Note that as long as we label the states, the order in which we write them does not matter! We now write these basis states with braces in the Bell basis using (3.59). We find

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= \frac{1}{2}(|\Phi_0\rangle_{AC} + |\Phi_3\rangle_{AC}) \alpha|+\rangle_B + \frac{1}{2}(|\Phi_1\rangle_{AC} - i|\Phi_2\rangle_{AC}) \beta|+\rangle_B \\ &\quad + \frac{1}{2}(|\Phi_1\rangle_{AC} + i|\Phi_2\rangle_{AC}) \alpha|-\rangle_B + \frac{1}{2}(|\Phi_0\rangle_{AC} - |\Phi_3\rangle_{AC}) \beta|-\rangle_B. \end{aligned} \quad (4.72)$$

Collecting the Bell states we find

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= \frac{1}{2}|\Phi_0\rangle_{AC} (\alpha|+\rangle_B + \beta|-\rangle_B) + \frac{1}{2}|\Phi_1\rangle_{AC} (\alpha|-\rangle_B + \beta|+\rangle_B) \\ &\quad + \frac{1}{2}|\Phi_2\rangle_{AC} (i\alpha|-\rangle_B - i\beta|+\rangle_B) + \frac{1}{2}|\Phi_3\rangle_{AC} (\alpha|+\rangle_B - \beta|-\rangle_B). \end{aligned} \quad (4.73)$$

We can then see that in fact we got

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= \frac{1}{2}|\Phi_0\rangle_{AC} \otimes |\Psi\rangle_B + \frac{1}{2}|\Phi_1\rangle_{AC} \otimes \sigma_1|\Psi\rangle_B \\ &\quad + \frac{1}{2}|\Phi_2\rangle_{AC} \otimes \sigma_2|\Psi\rangle_B + \frac{1}{2}|\Phi_3\rangle_{AC} \otimes \sigma_3|\Psi\rangle_B. \end{aligned} \quad (4.74)$$

The above right-hand side allows us to understand what happens when Alice measures the state of (A, C) in the Bell basis. If she measures:

- $|\Phi_0\rangle_{AC}$, then the B state becomes $|\Psi\rangle_B$,
- $|\Phi_1\rangle_{AC}$, then the B state becomes $\sigma_1|\Psi\rangle_B$,
- $|\Phi_2\rangle_{AC}$, then the B state becomes $\sigma_2|\Psi\rangle_B$,
- $|\Phi_3\rangle_{AC}$, then the B state becomes $\sigma_3|\Psi\rangle_B$.

If Alice got $|\Phi_0\rangle_{AC}$ then Bob is in the possession of the teleported state and has to do nothing. If Alice gets $|\Phi_i\rangle_{AC}$, Bob's particle is goes into the state $\sigma_i|\Psi\rangle_B$. Bob applies the i -th box, which multiplies his state by σ_i giving him the desired state $|\Psi\rangle_B$. The teleporting is thus complete!

Note that Alice is left with one of the Bell states $|\Phi_\mu\rangle_{AC}$ which has no information whatsoever about the constants α and β that defined the state to be teleported. Thus the process did not create a copy of the state. The original state is destroyed in the process of teleportation.

It is noteworthy that all the mathematical work above led to the key result (4.74), which is neatly summarized as the following identity valid for arbitrary states $|\Psi\rangle$:

$$|\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C = \frac{1}{2} \sum_{i=0}^3 |\Phi_i\rangle_{AC} \otimes \sigma_i|\Psi\rangle_B. \quad (4.75)$$

This is an identity for a state of three particles. It expresses the tensor product of an entangled state of the first two particles, times a third, as a sum of products that involve entangled states of the first and third particle times a state of the second particle.

5 EPR and Bell Inequalities

In this section we begin by studying some properties of the singlet state of two particles of spin-1/2. We then turn to the claims of Einstein, Podolsky, and Rosen (EPR) concerning quantum mechanics. Finally, we discuss the so-called Bell inequalities that would follow if EPR were right. Of course, quantum mechanics violates these inequalities, and experiment shows that the inequalities are indeed violated. EPR were wrong.

We have been talking about the singlet state of two spin-1/2 particles. This state emerges, for example, in particle decays. The neutral η_0 meson (of rest mass 547 MeV) sometimes decays into two oppositely charged muons

$$\eta_0 \rightarrow \mu^+ + \mu^- . \quad (5.1)$$

The meson is a spinless particle and being at rest has zero orbital angular momentum. As a result it has zero total angular momentum. As it decays, the final state of the two muons must have zero total angular momentum as well. If the state of the two muons has zero orbital angular momentum, it must also have zero total spin angular momentum. The two muons flying away from each other with zero orbital angular momentum are in a singlet state. This state takes the form

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|+\rangle_1|-\rangle_2 - |-\rangle_1|+\rangle_2) . \quad (5.2)$$

This singlet state is rotational invariant and therefore it is actually the same for whatever choice of direction \mathbf{n} to define a basis of spin states:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\mathbf{n};+\rangle_1|\mathbf{n};-\rangle_2 - |\mathbf{n};-\rangle_1|\mathbf{n};+\rangle_2) . \quad (5.3)$$

We now ask: In this singlet, what is the probability $P(\mathbf{a}, \mathbf{b})$ that the first particle is in the state $|\mathbf{a}; +\rangle$ and the second particle is in the state $|\mathbf{b}; +\rangle$, with \mathbf{a} and \mathbf{b} two arbitrarily chosen unit vectors? To help ourselves, we write the singlet state using the first vector

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a};+\rangle_1|\mathbf{a};-\rangle_2 - |\mathbf{a};-\rangle_1|\mathbf{a};+\rangle_2) . \quad (5.4)$$

By definition, the probability we want is

$$P(\mathbf{a}, \mathbf{b}) = \left| {}_1\langle \mathbf{a}; +| {}_2\langle \mathbf{b}; +|\Psi\rangle \right|^2 \quad (5.5)$$

Only the first term in (5.4) contributes and we get

$$P(\mathbf{a}, \mathbf{b}) = \frac{1}{2}|\langle \mathbf{b}; +|\mathbf{a}; -\rangle|^2 \quad (5.6)$$

We recall that the overlap-squared between two spin states is given by the cosine-squared of half the angle in between them. Using figure 2 we see that the angle between \mathbf{b} and $-\mathbf{a}$ is $\pi - \theta_{ab}$, where θ_{ab} is the angle between \mathbf{b} and \mathbf{a} . Therefore

$$P(\mathbf{a}, \mathbf{b}) = \frac{1}{2}\cos^2\left(\frac{1}{2}(\pi - \theta_{ab})\right) \quad (5.7)$$

Our final result is therefore

$$P(\mathbf{a}, \mathbf{b}) = \frac{1}{2} \sin^2\left(\frac{1}{2}\theta_{ab}\right). \quad (5.8)$$

As a simple consistency check, if $\mathbf{b} = -\mathbf{a}$ then $\theta_{ab} = \pi$ and $P(\mathbf{a}, -\mathbf{a}) = 1/2$ which is what we expect.

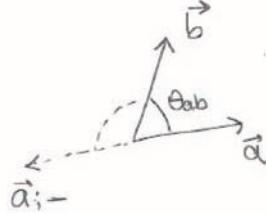


Figure 2: Directions associated with the vectors \mathbf{a} and \mathbf{b} .

If we measure about orthogonal vectors, like the unit vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ we get

$$P(\hat{\mathbf{z}}, \hat{\mathbf{x}}) = \frac{1}{2} \sin^2 45^\circ = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \quad (5.9)$$

The key statement of Einstein, Podolsky and Rosen (EPR) is the claim for **local realism**. This is posed as two properties of measurement:

1. The result of a measurement corresponds to some element of reality. If a measurement of an observable gives a value, that value was a property of the state.
2. The result of a measurement at one point cannot depend on whatever action takes place at a far away point at the same time.

Both properties seem quite plausible at first thought. The first, we are by now accustomed, is violated in Quantum Mechanics, where measurement involves collapse of the wavefunction, so that the result was not pre-ordained and does not correspond to a unequivocal property of the system. The violation of the second is perhaps equally disturbing, given our intuition that simultaneous spatially separated events can't affect each other. There is something non-local about quantum mechanics.

According to EPR the so called entangled pairs are just pairs of particles that have definite spins. They point out that the results of quantum mechanical measurements are reproduced if our large ensemble of pairs has the following distribution of states:

- In 50% of pairs, particle 1 has spin along $\hat{\mathbf{z}}$ and particle 2 has spin along $-\hat{\mathbf{z}}$,
- In 50% of pairs, particle 1 has spin along $-\hat{\mathbf{z}}$ and particle 2 has spin along $\hat{\mathbf{z}}$.

This would explain the perfect correlations and is consistent, for example, with $P(\hat{\mathbf{z}}, -\hat{\mathbf{z}}) = 1/2$, which we obtained quantum mechanically.

The challenge for EPR is to keep reproducing the results of more complicated measurements. Suppose each of the two observers can measure spin along two possible axes: the x and z axes. They measure once, in any of these two directions. EPR then state that in any pair each particle has a definite state of spin in these two directions. For example, a particle of type $(\hat{\mathbf{z}}, -\hat{\mathbf{x}})$ is one that if measured along z gives a plus $\hbar/2$ and if measured along x gives $-\hbar/2$. We do not do simultaneous measurements or subsequent measurements on each particle. EPR then claim that the observed quantum mechanical results are matched if our ensemble of pairs have the following properties

- 25% of pairs have particle 1 in $(\hat{\mathbf{z}}, \hat{\mathbf{x}})$ and particle 2 in $(-\hat{\mathbf{z}}, -\hat{\mathbf{x}})$
- 25% of pairs have particle 1 in $(\hat{\mathbf{z}}, -\hat{\mathbf{x}})$ and particle 2 in $(-\hat{\mathbf{z}}, \hat{\mathbf{x}})$
- 25% of pairs have particle 1 in $(-\hat{\mathbf{z}}, \hat{\mathbf{x}})$ and particle 2 in $(\hat{\mathbf{z}}, -\hat{\mathbf{x}})$
- 25% of pairs have particle 1 in $(-\hat{\mathbf{z}}, -\hat{\mathbf{x}})$ and particle 2 in $(\hat{\mathbf{z}}, \hat{\mathbf{x}})$

First note the complete correlations: particles one and two have opposite spins in each possible direction. This is, of course, needed to match the quantum mechanical singlets. We can ask what is $P(\hat{\mathbf{z}}, -\hat{\mathbf{z}})$, the probability that particle one is along $\hat{\mathbf{z}}$ and particle two along $-\hat{\mathbf{z}}$. The first two cases above apply, and thus this probability is $1/2$, consistent with quantum mechanics. We can also ask for $P(\hat{\mathbf{z}}, \hat{\mathbf{x}})$. This time only the second case applies giving us a probability of $1/4$ as we obtained earlier in (5.9). The quantum mechanical answers indeed arise.

The insight of Bell was that with Stern-Gerlach apparatuses that could measure in three directions one gets in trouble. Suppose each observer can measure along any one of the three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$. Again, each particle is just measured once. Let us assume that we have a large number N of pairs that, following EPR, contain particles with well-defined spins on these three directions. A particle of type $(\mathbf{a}, -\mathbf{b}, \mathbf{c})$, for example, if measured along \mathbf{a} would give $\hbar/2$, if measured along \mathbf{b} would give $-\hbar/2$ and if measured along \mathbf{c} would give $\hbar/2$. The following distribution is given:

Populations	Particle 1	Particle 2
N_1	$(\mathbf{a}, \mathbf{b}, \mathbf{c})$	$(-\mathbf{a}, -\mathbf{b}, -\mathbf{c})$
N_2	$(\mathbf{a}, \mathbf{b}, -\mathbf{c})$	$(-\mathbf{a}, -\mathbf{b}, \mathbf{c})$
N_3	$(\mathbf{a}, -\mathbf{b}, \mathbf{c})$	$(-\mathbf{a}, \mathbf{b}, -\mathbf{c})$
N_4	$(\mathbf{a}, -\mathbf{b}, -\mathbf{c})$	$(-\mathbf{a}, \mathbf{b}, \mathbf{c})$
N_5	$(-\mathbf{a}, \mathbf{b}, \mathbf{c})$	$(\mathbf{a}, -\mathbf{b}, -\mathbf{c})$
N_6	$(-\mathbf{a}, \mathbf{b}, -\mathbf{c})$	$(\mathbf{a}, -\mathbf{b}, \mathbf{c})$
N_7	$(-\mathbf{a}, -\mathbf{b}, \mathbf{c})$	$(\mathbf{a}, \mathbf{b}, -\mathbf{c})$
N_8	$(-\mathbf{a}, -\mathbf{b}, -\mathbf{c})$	$(\mathbf{a}, \mathbf{b}, \mathbf{c})$

As required, all spins are correlated in particles one and two. We also have $N = \sum_{i=1}^8 N_i$. We now record the following probabilities that follow by inspection of the table:

$$P(\mathbf{a}, \mathbf{b}) = \frac{N_3 + N_4}{N}, \quad P(\mathbf{a}, \mathbf{c}) = \frac{N_2 + N_4}{N}, \quad P(\mathbf{c}, \mathbf{b}) = \frac{N_3 + N_7}{N}. \quad (5.10)$$

Consider now the trivially correct inequality:

$$N_3 + N_4 \leq N_3 + N_7 + N_2 + N_4, \quad (5.11)$$

that on account of (5.10) implies the **Bell inequality**

$$\boxed{P(\mathbf{a}, \mathbf{b}) \leq P(\mathbf{a}, \mathbf{c}) + P(\mathbf{c}, \mathbf{b}).} \quad (5.12)$$

If true quantum mechanically, given (5.8) we would have

$$\frac{1}{2} \sin^2 \frac{1}{2}\theta_{ab} \leq \frac{1}{2} \sin^2 \frac{1}{2}\theta_{ac} + \frac{1}{2} \sin^2 \frac{1}{2}\theta_{cb}. \quad (5.13)$$

But this is violated for many choices of angles. Take, for example, the planar configuration in Fig. 3:

$$\theta_{ab} = 2\theta, \quad \theta_{ac} = \theta_{cb} = \theta. \quad (5.14)$$

For this situation, the inequality becomes

$$\frac{1}{2} \sin^2 \theta \leq \sin^2 \frac{1}{2}\theta. \quad (5.15)$$

This fails for sufficiently small θ : $\frac{1}{2}\theta^2 \leq \frac{\theta^2}{4}$ is just plain wrong. In fact, the inequality goes wrong for any $\theta < \frac{\pi}{2}$. Experimental results have confirmed that Bell inequalities are violated and thus the original claim of local realism by EPR is wrong.

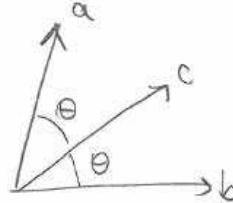


Figure 3: Special configuration for vectors \mathbf{a} , \mathbf{b} and \mathbf{c} .

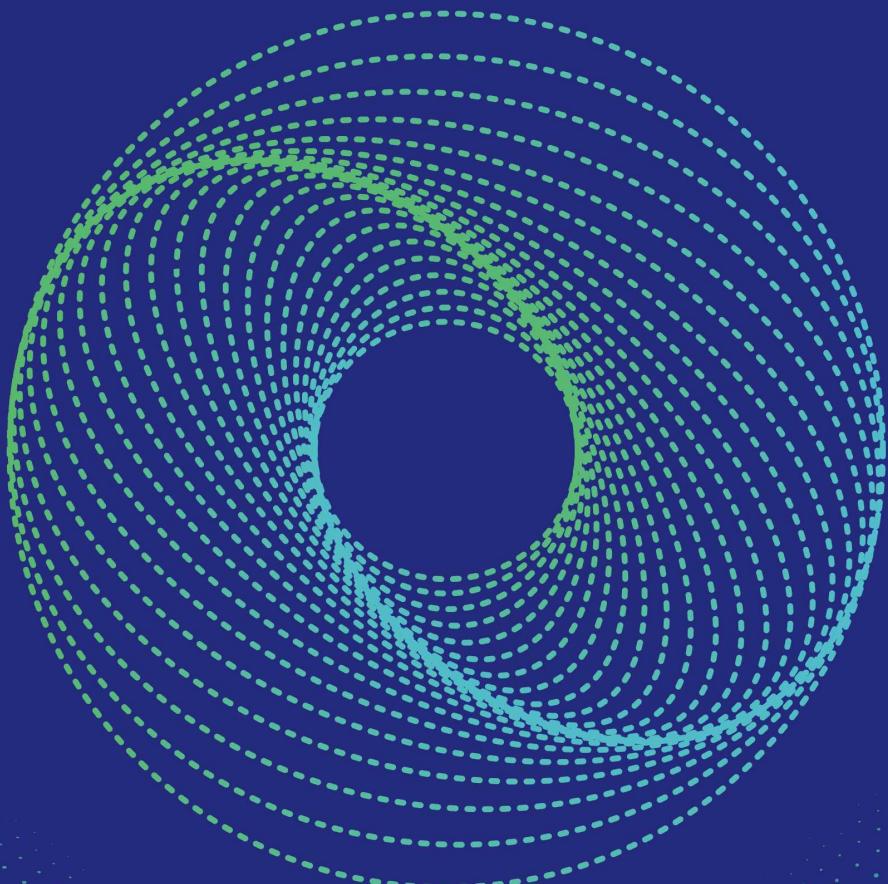
MIT OpenCourseWare
<http://ocw.mit.edu>

8.05 Quantum Physics II
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Quantum Computing & Information

A Scaffolding Approach



The Polaris QCI
Quantum Computing Series

Peter Y. Lee . Huiwen Ji . Ran Cheng

Peter Y. Lee : Earned a Ph.D. in E.E. from Princeton, specializing in quantum nanostructures and the fractional quantum Hall effect. Post-academia, he joined Bell Labs, contributing to photonics and securing 20+ patents. He brings extensive teaching experience and is now a faculty member at Fei Tian College, NY.

Huiwen Ji : Holds a Ph.D. in Chemistry from Princeton, with a focus on solid-state chemistry tied to quantum properties. Her research spans quantum physics to materials chemistry. With roles at UC Berkeley, and Lawrence Berkeley National Lab, she's received awards like the NSF CAREER Award. Currently a faculty member at the University of Utah.

Ran Cheng : Obtained a Ph.D. in Physics from University of Texas at Austin, specializing in condensed matter theory with a focus on spintronics and magnetism. He was a postdoc at Carnegie Mellon University. He is currently a faculty member at University of California, Riverside, where he received the NSF CAREER and DoD MURI awards.

"This impressive book covers the burgeoning field of quantum information, bridging the fundamentals of quantum mechanics and its present and future applications in secure communication and quantum computing. The author's approach is rigorous - including all the necessary linear algebra - while the book is highly readable and accessible. It will benefit a wide range of audiences with different backgrounds, from undergraduate students learning quantum mechanics to experts who want a deep understanding of quantum information protocols."

— Andrew Kent

Professor of Physics, Director of the Center for Quantum Phenomena, New York University

"Quantum computing is poised to be one of the first major technological developments of the 21st century. This book assumes a student has a solid background in quantum mechanics, which allows it to introduce the broad field of quantum information and computing in depth. At the same time it covers important topics from multiple angles, which is invaluable in guiding students who are first learning the material. It will serve well both for teaching and as a reference."

— Stephen Lyon

Professor of Elec. and Comp. Engineering, Princeton Quantum Initiative, Princeton University

"Quantum Computing is definitely going to impact our future lives. This book adheres to a pedagogical methodology that balances theoretical rigor with accessibility. The scaffolding approach that the authors use guides the reader through the learning journey. This makes the book not only academically rigorous but also effective as a teaching tool!"

— Robert J. Cava

Professor of Chemistry, Princeton Quantum Initiative, Princeton University



Quantum Computing and Information

A Scaffolding Approach

Peter Y. Lee
Huiwen Ji
Ran Cheng



Copyright © 2024 Polaris QCI Publishing

This work is subject to copyright. All rights are reserved. No part of this publication may be reproduced, reprinted, translated, uploaded to electronic storage systems, or transmitted, in any form or by any means, electronic, mechanical, or otherwise, without the prior written permission of the publisher. Circumventing digital rights management, unauthorized distribution, sharing, or sale of this publication is a violation of law and is subject to criminal prosecution.

ISBN 978-1-961-88000-9 (ebook)

ISBN 978-1-961-88001-6 (paperback, b/w)

ISBN 978-1-961-88002-3 (hardcover, b/w)

ISBN 978-1-961-88003-0 (paperback, color)

ISBN 978-1-961-88004-7 (hardcover, color)

Library of Congress Control Number (LCCN) 2024901045

First edition, March 2024

This document is typeset using L^AT_EX.

Quantum circuit drawings are created using the yquant package from
<https://github.com/projekter/yquant>.

Publisher website: <https://polarisqci.com>

About This Book

In a world where quantum computing stands at the crossroads of computation and quantum mechanics, *Quantum Computing and Information: A Scaffolding Approach* offers a meticulously designed pathway for mastering this transformative technology. As part of an educational series, this book serves as a comprehensive resource for beginning graduate students, senior undergraduates, and anyone invested in understanding the quantum computational landscape.

The book follows a “scaffolding approach,” inspired by pedagogical theories from Lev Vygotsky and Jerome Bruner, guiding readers through complex subject matter without overwhelming them. Through the gradual introduction of concepts, layered reinforcement, and practical exercises, the book facilitates deep learning. Employing ample illustrations, tables, and special boxes for highlights and key concepts, the text makes quantum computing accessible without diluting its intricacies.

Four major sections unfold a comprehensive learning journey: from understanding the basics of quantum systems, through the manipulation of these systems with quantum gates, to the fascinating phenomenon of entanglement, and finally, to essential quantum algorithms, error correction techniques, and quantum information theory.

Whether you are a novice to quantum computing or have some experience in the field, this book offers a structured and incremental approach to gaining a robust understanding. Get ready to embark on an enlightening voyage through the captivating realm of quantum computing.

About the Authors

Dr. Peter Y. Lee holds a Ph.D. in Electrical Engineering from Princeton University. His research at Princeton focused on quantum nanostructures, the fractional quantum Hall effect, and Wigner crystals. Following his academic tenure, he joined Bell Labs, making significant contributions to the fields of photonics and optical communications and securing over 20 patents. Dr. Lee's multifaceted expertise extends to educational settings; he has a rich history of teaching, academic program oversight, and computer programming. Dr. Lee is currently on the faculty of Fei Tian College, New York.

Dr. Huiwen Ji earned her Ph.D. in Chemistry at Princeton University, where she specialized in the solid-state chemistry of binary and ternary chalcogenides, a field intricately tied to quantum properties and topological surface states. This rigorous academic background laid the foundation for her subsequent research endeavors, blending quantum physics, materials chemistry, and structure-property relationships in solid-state functional materials. In her roles as a Postdoctoral Scholar at the University of California, Berkeley, and a Research Scientist at Lawrence Berkeley National Lab, she further delved into the nuances of advanced material science. Recognized for her significant contributions, Dr. Ji has received accolades such as the ACS PRF Doctoral New Investigator Award and the NSF CAREER Award. She currently serves as a faculty member at the University of Utah.

Dr. Ran Cheng earned his Ph.D. in Physics from the University of Texas at Austin with a focus on theoretical condensed matter physics. After receiving his doctorate, he became a postdoctoral researcher at Carnegie Mellon University to further his inquiry into magnetic materials. He is now a faculty member at the University of California, Riverside, where he explores three core research domains: spintronics, topological materials, and quantum magnets. A recognized pioneer in the burgeoning field of antiferromagnetic spintronics, Dr. Cheng was honored with the NSF CAREER award and the DoD MURI award alongside a cadre of distinguished physicists.

Contents

Level Indicators

Unmarked content: Foundational, appropriate for all readers, including newcomers to quantum computing.

*: Typically an initial conceptual overview accessible to general readers, followed by math-intensive segments for a senior undergraduate or early graduate-level audience with a background in linear algebra. Assumes familiarity with unmarked content.

*: Advanced exploration, extending beyond the *-level topics.

Preface	viii
Reviews	xii
About Quantum Computing and Information	xiv

I

Qubits & Qudits: Foundations

1	Quantum Mechanics Through Photons	2
1.1	Introducing Quantum Mechanics	3
1.2	Understanding Photons	6
1.3	The Quantum State Postulate	9
1.4	The Quantum Observable Postulate	16
1.5	The Quantum Measurement Postulate	23
1.6	The Uncertainty Principle	32
1.7	* Further Readings on Quantum Mechanics	35
1.8	Topic Reviews	36
1.9	Summary and Conclusions	39
	Problem Set 1	40
2	Fundamentals of Spin Systems	41
2.1	Spin, Angular Momentum, and Magnetic Moment	42
2.2	Spin-1/2 States and Pauli Matrices	43

2.3	* General Spin State Representation	45
2.4	The Bloch Sphere	48
2.5	Spin Measurement	49
2.6	Summary and Conclusions	53
	Problem Set 2	54
3	A Framework for Qubits and Qudits	56
3.1	Physical Qubit Systems	57
3.2	Qubit and Qudit States	57
3.3	Change of Basis	62
3.4	* General Formulation of Quantum Measurement	68
3.5	* Application to Quantum State Tomography	77
3.6	Summary and Conclusions	78
	Problem Set 3	80
4	Dynamics of Quantum Systems	82
4.1	The Evolution Postulate of Quantum Mechanics	83
4.2	* The Schrödinger Equation	84
4.3	Stationary Nature of Energy Eigenstates	88
4.4	Universal Quantum Computing and Annealing	91
4.5	* Larmor Precession and Rabi Oscillations	93
4.6	* Further Exploration	99
4.7	* Deferred Proofs	101
4.8	Summary and Conclusions	103
	Problem Set 4	104

II

Quantum Gates & Elementary Circuits

5	Single-Qubit Quantum Gates	107
5.1	Quantum Versus Classical Logic Gates	108
5.2	Common Single-Qubit Gates	111
5.3	From Gate Sequences to Quantum Circuits	119
5.4	Quantum Random Number Generator	122
5.5	The BB84 Quantum Key Distribution (QKD) Protocol	124
5.6	The Quantum Coin Game	131
5.7	* The No-Cloning Theorem: Proof Outline	134
5.8	Summary and Conclusions	135
	Problem Set 5	136
6	Multi-Qubit Systems	138
6.1	Systems of Two Qubits	139

6.2	* Measurements of Two-Qubit Systems	143
6.3	Multi-Qubit System States	151
6.4	* Measurements of Multi-Qubit Systems	157
6.5	* Time Evolution of Multi-Qubit States	162
6.6	Summary and Conclusions	166
	Problem Set 6	167
7	Multi-Qubit Quantum Gates	170
7.1	Common Multi-Qubit Gates	171
7.2	Universal Sets of Qubit Gates	179
7.3	Boolean Representation of Quantum Gates	182
7.4	Equivalent Gate Sequences	186
7.5	* Exploratory Topics	195
7.6	Summary and Conclusions	197
	Problem Set 7	198

III

Quantum Entanglement

8	Bell States	201
8.1	Maximally Entangled States	202
8.2	* Bell Basis	203
8.3	Bell State Creation	205
8.4	Bell Measurement	208
8.5	Bell State Conversion	209
8.6	* Deferred Proofs	211
8.7	* Generalization: GHZ States	212
8.8	Summary and Conclusions	213
	Problem Set 8	214
9	Entanglement and Bell Inequalities	217
9.1	Classical Correlation vs. Quantum Entanglement	218
9.2	The EPR Paradox	221
9.3	Bell Inequalities	223
9.4	Bell-CHSH Inequality with Classical Correlation	225
9.5	Bell-CHSH Inequality with Quantum Entanglement	227
9.6	Experimental Verification	230
9.7	The No-Communication Theorem	234
9.8	* Derivation of Bell-CHSH Quantity for Bell States	235
9.9	* Further Exploration	240
9.10	Summary and Conclusions	241

Problem Set 9	242
-------------------------	-----

10 Key Applications of Entanglement	245
10.1 Review of Preliminaries	246
10.2 Superdense Coding	248
10.3 Quantum Teleportation	251
10.4 Entanglement Swapping	260
10.5 Quantum Gate Teleportation	264
10.6 E91 Quantum Key Distribution Protocol	277
10.7 Summary and Conclusions	282
Problem Set 10	283

IV

Quantum Computation & Information

11 Quantum Algorithms: A Sampler	288
11.1 The Deutsch-Jozsa Algorithm	290
11.2 QUBO, VQE, QAOA, and AQC	298
11.3 Quantum Bomb and Quantum Money	315
11.4 Summary and Conclusions	328
Problem Set 11	329
12 Quantum Error Correction: A Primer	333
12.1 Preliminary Concepts	334
12.2 <i>* Mixed States, Density Operators, and CPTP Maps</i>	338
12.3 Error Mechanisms in Quantum Computing	361
12.4 Introduction to Error Correction Codes	376
12.5 Summary and Conclusions	391
Problem Set 12	392
13 Fundamentals of Quantum Information	396
13.1 Quantum Probability Essentials	399
13.2 Quantum Entropy and Information	417
13.3 <i>* Core Theorems in Quantum Information</i>	428
13.4 Further Exploration	435
13.5 Summary and Conclusions	435
Problem Set 13	436

Essential Mathematics: Quick References	442
A Complex Numbers	442
B Trigonometry	443
C Linear Algebra for QCI	445
D Pauli Matrices	452
Bibliography	455
List of Figures	464
List of Tables	465
Index	466
Journey Forward	472

Preface

This book serves as a part of a series of textbooks initially crafted for the Master of Science in Quantum Computing Program at Fei Tian College, Middletown, New York. The series aspires to offer a pedagogically sound, systematic approach to teaching and learning quantum computing. It includes the following titles:

- Mathematical Foundations of Quantum Computing
- Quantum Computing and Information: A Scaffolding Approach (current book)
- Quantum Algorithms and Applications: A Scaffolding Approach

While each book functions as a standalone guide to its respective topic, collectively they furnish a comprehensive understanding of quantum computing.

Designed for beginning graduate students and senior undergraduates, this book also includes markers to aid both entry-level and more advanced readers.

Quantum Computing and Information (QCI) is a complex discipline, comprising an intricate web of knowledge that spans advanced mathematics, quantum mechanics, and sophisticated algorithms. Navigating this multidimensional landscape requires an approach to learning and teaching that acknowledges the inherently linear nature of reading and lectures, while also addressing the multi-faceted structure of the subject matter.

Effective teaching—and, by extension, effective learning—is not merely the transmission of information but a dynamic process of constructing understanding within a cognitive framework. This involves connecting new knowledge with existing cognitive structures, akin to weaving new threads into an ever-expanding web of understanding in the brain.

The scaffolding approach, central to this text, is inspired by educational theories such as Lev Vygotsky's Zone of Proximal Development and Jerome Bruner's instructional strategies, refined through the authors' extensive experience in academia. It recognizes that learning in such a complex field as QCI involves navigating a path that gradually ascends from foundational concepts to advanced applications.

This book operationalizes the scaffolding approach through several key strategies:

- Progressive Introduction of Concepts: The content is structured to ensure learners are adequately prepared for each new concept, mirroring the idea of assembling a multidimensional puzzle piece by piece.
- Grounding in Simple Examples: Initially, straightforward examples and intuitive explanations are used to anchor new topics, which gradually evolve into more abstract and generalized concepts.

- Spiral Learning: Key ideas are revisited from multiple perspectives and in varying contexts, embodying the spiral nature of learning where concepts are deepened and expanded upon with each iteration.
- Active Engagement: Exercises and problems are integrated to promote deep learning through practical application, allowing learners to apply and test their understanding in meaningful ways.
- Cognitive Load Management: The book is designed with clarity in mind, featuring abundant illustrations and tables to complement the text, thereby reducing the cognitive load and demystifying complex mathematical and theoretical constructs.

By incorporating these elements, the scaffolding approach in this text not only imparts knowledge but also cultivates the skills necessary to navigate the multifaceted and interlinked domains of QCI. It reflects a sophisticated blend of the art and science of teaching, tailored to the unique challenges of this cutting-edge field.

Parts of the Book

The book is divided into four primary sections:

- Part I lays the groundwork for understanding quantum systems, commencing with simpler entities like photons and advancing to more complex systems such as qubits and qudits.
- Part II transitions from understanding individual quantum systems to manipulating them using quantum gates, acting as a bridge to more advanced quantum operations.
- Part III delves into the fascinating and essential quantum phenomenon of entanglement, starting with simple Bell States and expanding into its broader theory and applications.
- Part IV explores critical components of quantum computing and information: it offers a sampler of quantum algorithms, introduces the fundamentals of error correction, and delves into the principles of quantum information, forming a comprehensive overview of quantum computation and information.

Navigation Aids

To assist readers in navigating the book, the layout incorporates the following features:

Concept Box

This box is reserved for important concepts, postulates, and theorems.

Highlight Box

This box highlights corollaries, summaries, implications, and other key points.

Exercise 0.1 The exercises interspersed throughout the text serve to hone your skills and reinforce your understanding of the material discussed.

Each chapter concludes with a comprehensive Problem Set designed for thorough mastery. These problems are generally more challenging than the in-text exercises, providing an opportunity for deeper engagement with the content.



This indicator is used for tips, alerts, connections between concepts, and pieces of advice.

Info Box

This box provides supplementary context, associated concepts, and additional information.

Level Indicators

In order to accommodate learners with different academic backgrounds, some sections are marked by level indicators.

Unmarked content serves as the foundational basis for this subject and is appropriate for all readers, including those new to quantum computing.

Items marked with * typically begin with an introductory conceptual overview suitable for general readers, followed by math-intensive content designed for a senior undergraduate or early graduate-level audience with a background in linear algebra. Both components of * presuppose familiarity with unmarked foundational content.

Items marked with ** target further exploration and delve into topics that extend beyond the material covered at the * level.

Acknowledgements

To Ms. Elsie He and Ms. Jing Hunter, we express our deep gratitude for your artistic talents in the design of our book cover. Your perceptive interpretation of the subject matter and steadfast commitment to crafting a cover that is both visually arresting and thematically profound are truly valued.

Our heartfelt gratitude goes to Ms. Nathalie Chiao and Ms. Jenny Chang. Your editorial expertise has significantly enhanced the quality, readability, and coherence of our manuscript. Your meticulous attention to detail and commitment to clear communication have been invaluable.

To our esteemed colleagues and students, Mr. Yiyong Huang, Dr. Jason Wang, Dr. Joseph Zhao, and Dr. James Yu, we are immensely grateful. Your intellectual curiosity, unwavering support, and insightful questions have been pivotal in motivating us and shaping the approach of this book. Your feedback has been crucial in ensuring we provide practical understanding alongside a clear educational progression.

We also wish to express our sincerest thanks to our knowledgeable reviewers. Your expert reviews and constructive suggestions have played a key role in enhancing

the rigor and accessibility of this text, enabling us to present complex concepts in an engaging and informative manner.

- Robert J. Cava, Professor of Chemistry, Princeton Quantum Initiative, Princeton University
- Andrew Kent, Professor of Physics, The Center for Quantum Phenomena, New York University
- Shuwang Li, Professor of Applied Mathematics, Illinois Institute of Technology
- Stephen A. Lyon, Professor of Electrical and Computer Engineering, Princeton Quantum Initiative, Princeton University
- Leonid Pryadko, Professor of Physics, University of California at Riverside

Finally, we are indebted to the wider quantum computing community, upon whose advancements we build. Our goal is to distill and share this knowledge, making it accessible to students and inspiring further exploration into this fascinating field. Special thanks to the following influential experts:

- Scott Aaronson, Professor of Theoretical Computer Science, University of Texas at Austin; director of its Quantum Information Center
- John Preskill, Feynman Professor of Theoretical Physics, California Institute of Technology
- Ryan O'Donnell, Professor of Computer Science, Carnegie Mellon University
- John Watrous, Technical Director for Education, IBM Quantum

Dedication

This book is dedicated to all the inquisitive minds and relentless spirits who believe in the power of learning and the unbounded possibilities of human intellect.

Reviews

“ Quantum Computing is definitely going to impact our future lives. This book adheres to a pedagogical methodology that balances theoretical rigor with accessibility. The scaffolding approach that the authors use guides the reader through the learning journey. This makes the book not only academically rigorous but also effective as a teaching tool.”

— Robert J. Cava, Professor of Chemistry
Princeton Quantum Initiative, Princeton University

“ This impressive book covers the burgeoning field of quantum information, bridging the fundamentals of quantum mechanics and its present and future applications in secure communication and quantum computing. The author’s approach is rigorous – including all the necessary linear algebra – while the book is highly readable and accessible. It will benefit a wide range of audiences with different backgrounds, from undergraduate students learning quantum mechanics to experts who want a deep understanding of quantum information protocols.”

— Andrew Kent, Professor of Physics
The Center for Quantum Phenomena, New York University

“ Quantum computing is poised to be one of the first major technological developments of the 21st century. This book assumes a student has a solid background in quantum mechanics, which allows it to introduce the broad field of quantum information and computing in depth. At the same time it covers important topics from multiple angles, which is invaluable in guiding students who are first learning the material. It will serve well both for teaching and as a reference.”

— Stephen Lyon, Professor of Electrical and Computer Engineering
Princeton Quantum Initiative, Princeton University

“This textbook is elegantly crafted, utilizing a unique ‘scaffolding approach’ to render complex topics in quantum computing easily comprehensible for newcomers to the field. It is invaluable for both educators and students of quantum computing.

Quantum Computing and Information: A Scaffolding Approach offers a comprehensive and insightful introduction to quantum computing. Targeted at upper-division undergraduates with a foundational grasp of linear algebra or first-year graduates, it serves as an excellent resource for a one-semester course. The authors employ a lucid and engaging style, ensuring that complex topics are accessible. Their original illustrations and tables, meticulously designed to complement the text, enhance comprehension. Additionally, the textbook provides both concise and detailed examples, aiding entry-level students in grasping fundamental concepts. A well-considered balance between straightforward exercises (to consolidate specific knowledge) and problems (to integrate a broader understanding) is maintained.”

— *Shuwang Li, Professor of Applied Mathematics
Illinois Institute of Technology*

About Quantum Computing and Information

Introduction

Quantum Computing and Information (QCI) herald a paradigm shift in computational and information sciences, leveraging quantum mechanical principles to tackle complex problems across diverse fields such as cryptography, finance, and material science. Quantum computing, a cornerstone of QCI, focuses on utilizing quantum mechanics for computation, encompassing the development of algorithms, processors, and software. Extending beyond computing, quantum information pertains to the processing, storage, and transmission of information, while quantum sensing utilizes quantum properties for precise measurements. As QCI transitions from theoretical exploration to practical applications, its potential to revolutionize industry and academia becomes increasingly apparent.

Potential and Challenges

Quantum computing promises unprecedented speed in performing calculations for certain problem types, far exceeding classical computing capabilities. Algorithms like Shor's and Grover's exemplify quantum computing's superiority in tasks such as integer factorization and database searching. Yet, this potential is tempered by challenges in scalability, quantum noise, and operational costs. As we venture into systems with 100 qubits and beyond, the inadequacy of classical computers to simulate quantum systems becomes evident, highlighting the quantum advantage.

The NISQ Era and Beyond

Currently, we find ourselves in the Noisy Intermediate-Scale Quantum (NISQ) era, characterized by quantum processors that hold the promise of computations beyond classical simulation yet lack comprehensive error correction capabilities. This era has spurred the development of hybrid quantum-classical algorithms to optimize the computational power of NISQ devices amid noise and limited qubit coherence.

Transitioning from the NISQ era, we are beginning to witness the emergence of Quantum Utility. This new phase is not merely quantified by the number of qubits but by the capacity for reliable, significant computations that address real-world problems, marking a critical step from academic inquiry to industrial application. Quantum utility signifies the commencement of quantum computing's tangible benefits, driven by technological advancements that address enterprise-level challenges.

Workforce and Education

The burgeoning interest in QCI has precipitated a growing demand for professionals versed in quantum technologies. In response, educational institutions are integrating quantum computing into their curricula, anticipating a rise in the quest for quantum talent. Accessible software platforms will likely play a key role in democratizing quantum computing education, akin to the role of machine learning platforms today.

Required Expertise

The interdisciplinary nature of QCI necessitates expertise in linear algebra, quantum mechanics, computer science, and aspects of electrical engineering, among others. Proficiency with quantum programming languages and platforms such as Qiskit, Cirq, Q#, or Braket becomes increasingly important. As the field evolves, so too will the requisite skill set.

Learning Pathway

For those new to the field, a structured learning approach is recommended. Starting with foundational concepts like linear algebra and quantum mechanics before progressing to practical applications in quantum algorithms and programming ensures a solid base. Keeping abreast of the latest developments through journals, conferences, and academic courses is essential in a rapidly advancing field.

Are You Ready?

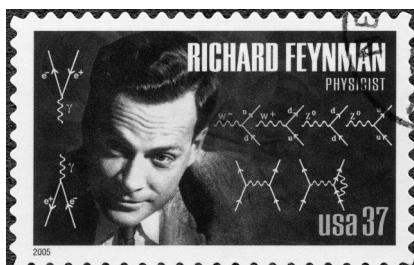
QCI represents not just a technological leap but a fundamental reimagining of computation and information processing. At this pivotal moment, the future of computing is poised for transformation, laden with challenges but rich in opportunity. The invitation is clear: embrace this cutting-edge technology, arm yourself with the necessary knowledge, and contribute to a future powered by the vast potential of QCI.

A Pioneering Vision for Quantum Computing

“ Nature isn’t classical, and if you want to make a simulation of nature, you’d better make it quantum mechanical, and it’s a wonderful problem, because it doesn’t look so easy.”

— Richard Feynman, 1981

First Conference on the Physics of Computation
Photo credit: Smithsonian Institution





Qubits & Qudits: Foundations

1	Quantum Mechanics Through Photons	2
2	Fundamentals of Spin Systems	41
3	A Framework for Qubits and Qudits	56
4	Dynamics of Quantum Systems	82

1. Quantum Mechanics Through Photons

Contents

1.1	Introducing Quantum Mechanics	3
1.1.1	A Brief History	3
1.1.2	Principles of Quantum Mechanics	4
1.1.3	Our Approach	5
1.2	Understanding Photons	6
1.2.1	What Are Photons?	6
1.2.2	Polarization of Photons	6
1.2.3	A Curious Light Polarization Experiment	8
1.3	The Quantum State Postulate	9
1.3.1	Classical Variables to Quantum Vectors	9
1.3.2	Rectilinear Polarization States	10
1.3.3	Basis States	11
1.3.4	Diagonal Polarization States	12
1.3.5	Circular Polarization States	14
1.3.6	General Polarization States	15
1.3.7	The Quantum State Postulate	16
1.4	The Quantum Observable Postulate	16
1.4.1	Macroscopic Quantities to Quantum Observables	17
1.4.2	Observables for Rectilinear Polarizations	19
1.4.3	Observables for Diagonal Polarizations	20
1.4.4	Observables for Circular Polarizations	20
1.4.5	Basis Equivalence and Dirac Notation	21
1.5	The Quantum Measurement Postulate	23
1.5.1	Measurement Probability	23
1.5.2	State Collapse	25
1.5.3	* Statistical Average	28
1.5.4	Quantum Interference	30
1.6	The Uncertainty Principle	32
1.6.1	Compatible versus Incompatible Observables	33
1.6.2	* The Uncertainty Inequality	34

1.7	* Further Readings on Quantum Mechanics	35
1.8	Topic Reviews	36
1.8.1	Review of Postulates 1-3	36
1.8.2	Review of Photons as Quantum Systems	37
1.9	Summary and Conclusions	39
	Problem Set 1	40

Quantum mechanics (QM) serves as the cornerstone for understanding the fundamental nature of our universe. It is the theoretical framework that governs the behavior of matter and energy on microscopic scales. Its principles have profound implications for emerging fields such as quantum computing, quantum communication, and quantum information. However, the abstract and counterintuitive nature of QM can present significant challenges for those unfamiliar with traditional physics and mathematics.

Photons offer a gateway into the world of quantum mechanics that is both accessible and illustrative. As the simplest system that embodies the core principles of QM, photons provide an opportunity to bypass the mathematical complexities often associated with this field. Photon phenomena can be easily visualized, enabling a more intuitive understanding. Moreover, photons play a crucial role as a key physical platform for quantum computing and quantum communication. This makes them not only a pedagogical tool but also a practical one, bridging theory and application.

The purpose of this chapter is to introduce the first three principles of Quantum Mechanics through the lens of photon behavior. The aim is to emphasize the basic concepts and principles essential to quantum computing, providing an easier learning path for readers who have not come from a physics background. By grounding the discussion in tangible photon phenomena, the chapter will seek to demystify QM, laying a solid foundation for further exploration and study in the interconnected fields of quantum computing, communication, and information.

1.1 Introducing Quantum Mechanics

1.1.1 A Brief History

At the turn of the 20th century, classical physics had successfully explained many macroscopic phenomena, including Newtonian mechanics, electromagnetism, thermodynamics, and more. Yet, certain physical observations, such as the photoelectric effect and the black body radiation problem, could not be explained within this classical framework.

Quantum mechanics emerged as a revolutionary theory that could account for these discrepancies. It became the fundamental theory describing matter, energy, and motion at atomic and subatomic scales. Technologies such as semiconductor

devices, lasers, and nuclear energy, and emerging fields like quantum computing, are rooted in quantum mechanics.

Key figures in the development of quantum mechanics include Planck, Einstein, Heisenberg, Bohr, Born, Schrödinger, and Dirac. Theories and notations named after these pioneers continue to be vital in current research and applications.

Quantum physics supersedes its classical counterparts, rendering classical physics as a special case of quantum mechanics. However, a full unification between quantum mechanics and general relativity remains an open question, with theories like superstring theory attempting to bridge this gap.

1.1.2 Principles of Quantum Mechanics

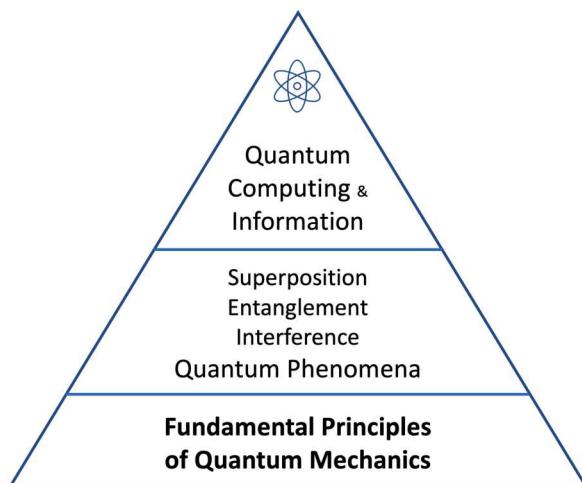


Figure 1.1: Role of Quantum Mechanics in Quantum Computing

As depicted in Fig. 1.1, the core components of quantum computing and information are grounded in specific quantum mechanical phenomena. Phenomena such as quantum superposition, entanglement, and interference are essential, but these complex behaviors are themselves derived from more fundamental principles of quantum mechanics.

Often referred to as postulates or axioms, these principles are the foundational pillars of quantum mechanics, analogous to how Newton's laws form the foundation of classical mechanics. Below is a brief introduction to the five principles, contextualized within the realm of quantum computing:

1. **Quantum States:** A quantum state is a complex vector that contains all the information about a quantum system. Quantum bits, or qubits, replace classical bits in quantum computing, and the state of a qubit is described as a superposition of the 0 and 1 states. This means a qubit can represent both states simultaneously, a feature that is unique to quantum systems. This principle enables parallel computation and has the potential to vastly increase computational efficiency for specific problems.

2. **Quantum Observables:** In quantum mechanics, observables are measurable physical quantities represented by Hermitian operators. In quantum computing, these operators extract specific outcomes from a system's quantum state during measurement.
3. **Quantum Measurements:** Quantum measurements facilitate information extraction from a quantum system, collapsing the state into one of the possible outcomes. This principle is essential in quantum algorithms, where measurements obtain the final result, and it is fundamental in quantum error correction.
4. **Quantum Evolution:** Quantum evolution governs the change in quantum states over time according to the Schrödinger equation. In quantum computing, quantum gates control this evolution, and algorithms are constructed using sequences of these gates.
5. **Quantum Composition:** Quantum composition involves building complex quantum systems from simpler ones. In quantum computing, this principle enables the creation of intricate quantum circuits from basic quantum gates, fostering the design of complex algorithms and enhancing the understanding of quantum computational processes.

These principles underpin the theoretical foundation of quantum computing, shaping the design, function, and comprehension of quantum algorithms, circuits, and systems. For those in the fields of quantum computing and quantum information, understanding these principles is vital, as they directly inform the practical aspects of computation and algorithm design.

In this chapter, we will focus on the first three principles, with the remaining two explored in subsequent chapters.

1.1.3 Our Approach

We strive to make Quantum Mechanics accessible to non-physics majors, tailored specifically for quantum computing and information (QCI), without sacrificing the rigor of core concepts. Our scaffolding approach unfolds the myriad of QM concepts gradually but systematically across three chapters (1-3) focused on photons, $\frac{1}{2}$ -spins, and qubits/qudits.

Traditional courses in quantum mechanics often commence with complex mathematical formulations of differential equations, applicable to a wide range of quantum systems. While this approach is mathematically rigorous, it does not align well with the specialized needs for understanding QCI.

In contrast, our approach uses linear algebra and emphasizes two-level quantum systems, particularly qubits, focusing on principles and techniques that are essential to quantum computing. By using tangible examples involving photons and electron spins, we render these abstract concepts more accessible. This methodology ensures an intuitive grasp of the subject matter without compromising academic rigor.

Furthermore, we acknowledge that QCI is an interdisciplinary field, bridging the domains of quantum physics, computer science, and mathematics. By situating the subject matter within this broader interdisciplinary context, we offer a more comprehensive perspective that extends beyond the scope of a conventional course

in quantum physics, thereby preparing the reader for the multifaceted landscape of QCI.

1.2 Understanding Photons

1.2.1 What Are Photons?

Photons are the particles of light. In fact, they constitute the particles of all electromagnetic waves, including radio waves, microwaves, visible light, X-rays, gamma rays, and more. You may wonder, “We have never heard that radio waves have photons; do they?” The answer is yes. In the case of radio waves, since their frequency is low and the wavelength is large, ranging from meters to 100 meters, we find ourselves in the classical regime where the quantum effects of the photons are not apparent.

However, quantum effects become crucial in the behavior of electromagnetic radiation under certain conditions, such as interactions at small scales (atomic and subatomic), high energy regimes, and strongly confined systems. Under these circumstances, classical electromagnetic theories fall short in describing phenomena like energy quantization and wave-particle duality. In fields like quantum computing and quantum communication, it’s not just that quantum effects are prominent; they are intentionally harnessed and manipulated, requiring a quantum treatment of light as well.

In the quantum regime, the energy of the electromagnetic waves is quantized, and the energy of each packet or photon depends on the frequency ($\nu = \omega/2\pi$), not on the amplitude, of the electromagnetic radiation:

$$E = h\nu = \hbar\omega \quad (1.1)$$

where h is the Planck constant, and $\hbar = h/2\pi$ is the reduced Planck constant.

In our everyday classical regime, an electromagnetic wave can be viewed as a flux of photons. For example, when looking at a computer screen, our eyes receive somewhere between 10^{14} to 10^{16} photons per second. However, in quantum computing and communication, we can transmit, receive, and measure a single photon at a time.

1.2.2 Polarization of Photons

In classical electromagnetism, polarization refers to the orientation of the electric field vector's oscillations in an electromagnetic wave. A polarized light wave consists of electric and magnetic fields oscillating in specific directions as the wave propagates. Linear polarization, where the electric field vector oscillates along a single axis, is common, but circular and elliptical polarization also exist. Various optical devices such as polarizers can manipulate or analyze classical light wave polarization.

In the quantum realm, polarization takes on a more abstract representation. The polarization of a photon corresponds to an intrinsic property related to its angular momentum. A photon is treated as a quantum particle, and its polarization is a two-dimensional complex vector, corresponding to a quantum state. Unlike the continuous oscillations of the electric field in the classical description, the quantum description deals with discrete outcomes. Quantum polarization can still be linear,

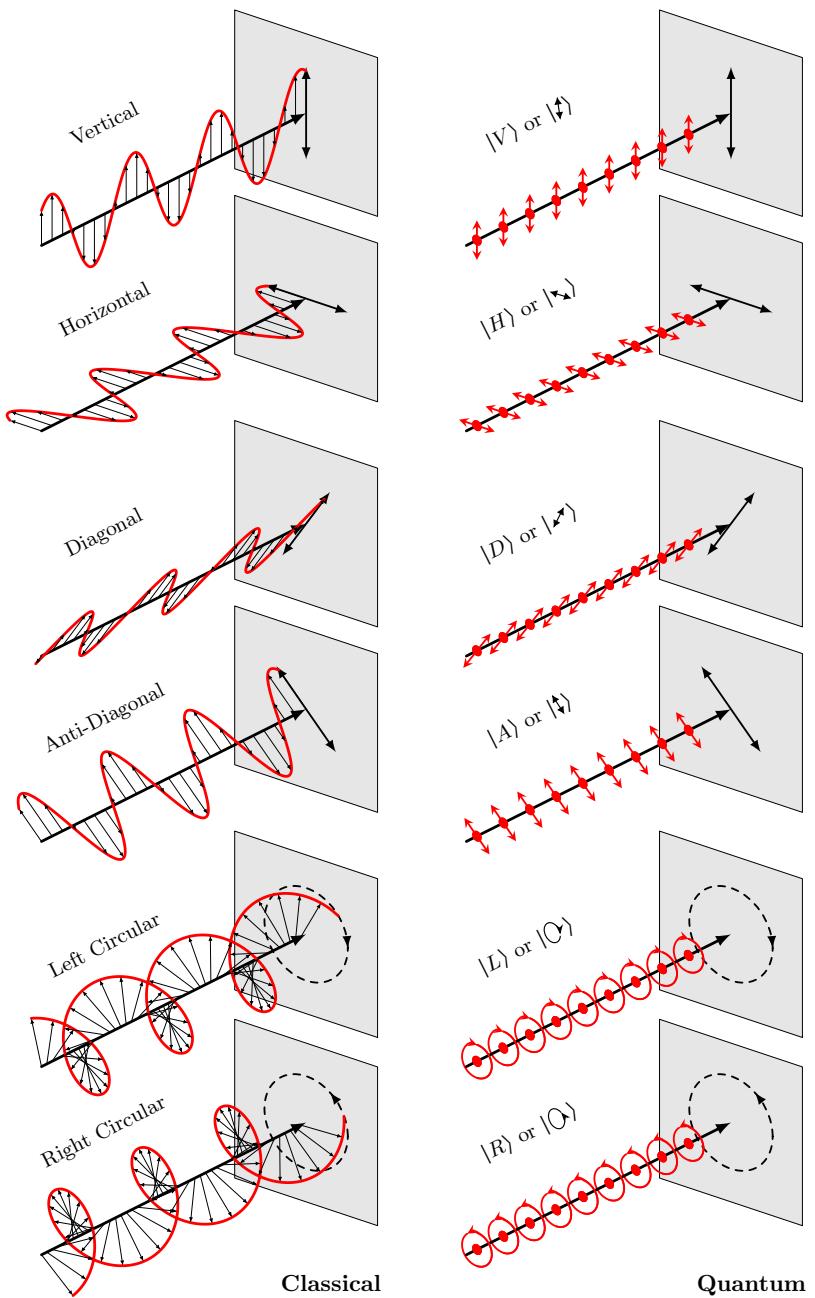


Figure 1.2: Light Polarization in Classical and Quantum Realms

Note: R and L in circular polarization refer to the rotation direction of the electric field vector as seen in the direction of propagation.

circular, or elliptical, and is vital in various quantum technologies such as quantum cryptography and quantum computing. Quantum polarization states can also be superposed and entangled, giving rise to distinctly quantum behaviors with no classical analogs.

Figure 1.2 illustrates light polarizations in the classical realm as waves of electric field vibration and as flux of photons in the quantum realm. The relationship between classical and quantum polarization descriptions lies in their conceptual connection and experimental consistency. While the classical view describes continuous wave properties, and the quantum view deals with particle-like attributes, polarization measurements in both realms yield consistent results. The quantum polarization state of a photon can be related to the corresponding classical wave's electric field orientation. This connection enables us to use photons as a model system to introduce the fundamental principles of quantum mechanics conveniently.

Table 1.1 lists key polarization states examined in this chapter, corresponding to Fig. 1.2. Additionally, elliptically polarized light is the most general state and can be considered as a linear combination of various orthogonal polarizations. Unpolarized light is a uniform statistical combination of these orthogonal states.

Category	Orthogonal States	Symbols
Rectilinear	Vertical	$ V\rangle$ or $ \uparrow\rangle$
	Horizontal	$ H\rangle$ or $ \leftrightarrow\rangle$
Diagonal	Diagonal (45°)	$ D\rangle$ or $ \swarrow\rangle$
	Anti-Diagonal (135°)	$ A\rangle$ or $ \nwarrow\rangle$
Circular	Right(handed)	$ R\rangle$ or $ \odot\rangle$
	Left(handed)	$ L\rangle$ or $ \ominus\rangle$

Table 1.1: Key Photon Polarization States

1.2.3 A Curious Light Polarization Experiment

Figure 1.3 depicts an experiment involving light polarization. In the background, there are some unpolarized light sources. Two polarizers are placed in front of the light sources, with one (H) oriented horizontally and the other (V) vertically. When observing either of these polarizers alone, we notice that about half of the light is transmitted. However, in the region where the two polarizers overlap, the light is entirely blocked.

A third polarizer (D), oriented at 45° , is then inserted between polarizers H and V. Intuitively, one might expect this third polarizer to block even more light, and that all the light would still be completely blocked where the three polarizers overlap. This intuition holds in the regions where H and D or V and D overlap, with about 25% of light being transmitted. Surprisingly, however, in the region where all three polarizers overlap, approximately 12.5% of the light is now transmitted. This situation does not occur if the third polarizer is placed on top or beneath H and V or if it is also oriented horizontally or vertically.

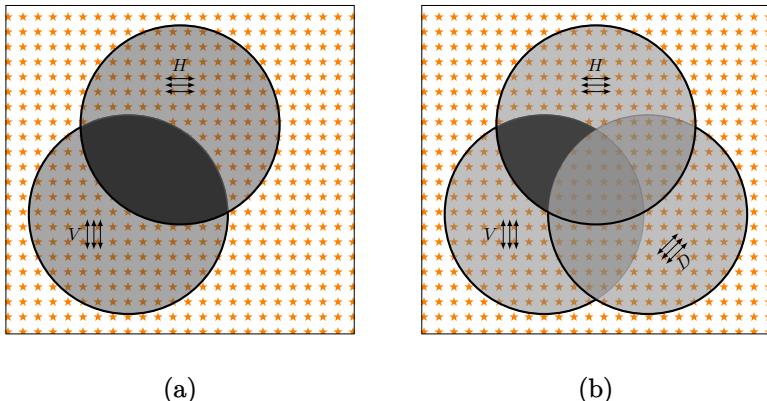


Figure 1.3: A Light Polarization Experiment

This intriguing behavior can be understood within the classical regime using traditional electromagnetic theory. In fact, you can demonstrate the experiment using Polaroid films as polarizers.

In the quantum regime, where it is possible to experiment with one photon at a time, the result is exactly the same. This raises a puzzling question: if H and V polarizations alone can block all photons, why do some photons pass after inserting another polarizer in between?

Quantum mechanics provides a comprehensive explanation for this experiment, even quantitatively. We will proceed to explain this experiment in terms of quantum mechanics, but first, we need to introduce some fundamental principles of quantum mechanics.

1.3 The Quantum State Postulate

1.3.1 Classical Variables to Quantum Vectors

The foundational principle of quantum mechanics is the principle of the quantum state. To understand this, let's first contrast it with classical physics, where the state of systems is described by recognizable physical quantities.

In a classical mechanical system, such as a ball moving through space, the state is characterized by position, velocity, and acceleration. In the case of a classical electromagnetic wave, the state is described by the **E** and **B** vectors, representing the electric and magnetic field strengths.

Quantum systems deviate substantially from their classical counterparts. Instead of conventional quantities like position or velocity, the state of a quantum system is represented by a special vector known as the state vector or wave function.

For a two-level quantum system like photon polarization, the state vector is two-dimensional with components α and β , both of which are generally complex numbers. Unlike classical vectors, where components are real numbers, the squared magnitudes of α and β relate to the probability of specific measurements. This

probabilistic nature requires the state vector to be normalized, meaning the sum of the squares of these components equals one.

Postulate 1: Quantum State

(Simplified version) The state of a quantum system is completely described by a normalized complex vector.

For instance, in a two-level system such as photon polarization, the quantum state is a two-dimensional vector, typically expressed as

$$|\psi\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (1.2)$$

To declare $|\psi\rangle$ as normalized means:

$$\langle\psi|\psi\rangle = |\alpha|^2 + |\beta|^2 = 1. \quad (1.3)$$

While photon polarization involves a two-dimensional state vector, the dimensionality of state vectors in other quantum systems can be much larger, even extending to infinity.



The vector representing a quantum state is often referred to as the *state vector*, or simply, the *state*. In some contexts, especially in infinite-dimensional space, it is also called the *wavefunction*.



Throughout this text, we use the *Dirac notation* for vectors and operators, offering a concise and powerful way to represent quantum states and operations. In this notation, $|\psi\rangle$ represents a column vector (where ψ is a label), while $\langle\psi|$ is its adjoint (the complex conjugate transpose, $\langle\psi| \equiv |\psi\rangle^\dagger$). The inner product between $|\phi\rangle$ and $|\psi\rangle$ is expressed as $\langle\phi|\psi\rangle$. For further resources on Dirac notation and linear algebra, please refer to Appendix C.

Exercise 1.1 Warm-up exercise on state vectors using Dirac notation.

Given $|a\rangle = \begin{bmatrix} \sqrt{2} + i\sqrt{3} \\ \sqrt{2} - i\sqrt{3} \end{bmatrix}$,

- (a) find $\langle a|$,
- (b) normalize $|a\rangle$, and
- (c) find a normalized vector $|b\rangle$ such that $\langle a|b\rangle = 0$.

1.3.2 Rectilinear Polarization States

Let's first take vertical and horizontal Rectilinear polarizations as examples to elucidate the Quantum State Postulate. We will begin by selecting vertical polarization as our reference state and assigning it the simplest normalized two-dimensional vector:

$$|V\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (1.4)$$



The content inside $|\cdot\rangle$ serves merely as a label. For instance, vertical polarization can be expressed in various ways such as $|\text{vertical}\rangle$, $|\Downarrow\rangle$, or $|V\rangle$.

What is the state vector for horizontal polarization, denoted as $|H\rangle$? Experimental evidence (refer to § 1.2.3) shows that vertically polarized photons will not pass through a horizontal polarizer, indicating that vertical and horizontal polarizations are mutually exclusive.

In quantum terms, mutually exclusive states correspond to disjoint measurement outcomes. Such states cannot coexist; if a system is in one of these states, it cannot simultaneously be in another. This exclusivity leads to the orthogonality of the states, meaning their inner product is zero: $\langle V|H \rangle = 0$.

Implication 1.1: Orthogonal States

Mutually exclusive quantum states are orthogonal; that is, their inner products are 0. Since these states are both normalized and orthogonal, they are referred to as *orthonormal*.

Given this orthogonality, the simplest normalized vector representing $|H\rangle$ and orthogonal to $|V\rangle$ is

$$|H\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (1.5)$$

1.3.3 Basis States

We recognize in linear algebra that a complete set of orthonormal vectors can serve as a basis. In our context, the vectors $|V\rangle$ and $|H\rangle$ are orthonormal. Additionally, since these vectors are linearly independent, they span the entire two-dimensional space, thus forming a complete set. The completeness ensures that any possible vector in the space can be represented as a linear combination of these vectors, allowing them to function as basis states. This leads to the expression of any polarization state, $|P\rangle$, as

$$|P\rangle = \alpha |V\rangle + \beta |H\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (1.6)$$

where α and β are complex numbers satisfying the normalization condition $|\alpha|^2 + |\beta|^2 = 1$.

Implication 1.2: Superposition States

A general quantum state exists as a linear superposition of basis states.

In forthcoming sections, we will elaborate on this idea using examples of diagonal and circular polarizations. However, before delving into those specifics, let's discuss the nature of our chosen basis states further.

We have designated the set $\{|V\rangle, |H\rangle\}$ as our *standard basis* for the system. In the parlance of quantum computing, this standard basis is also commonly referred to as the *computational basis*, represented by $\{|0\rangle, |1\rangle\}$. For our photon polarization system, the equivalence between these sets is formalized as

$$|0\rangle \equiv |V\rangle, \quad |1\rangle \equiv |H\rangle. \quad (1.7)$$

Therefore, Eq. 1.6 can be reformulated in terms of the computational basis as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \quad (1.8)$$



Equations 1.6 and 1.8 can be likened to the representation of a two-dimensional real vector $\mathbf{v} = a\hat{x} + b\hat{y}$ for easier visualization.

1.3.4 Diagonal Polarization States

Let's look at diagonal polarization. We said earlier that photon polarization is a system that's very easy to visualize in quantum mechanics. Let's illustrate why we can, in fact, use regular vectors to represent the polarization states. As shown in Fig. 1.4, the vertical polarization $|V\rangle$ is along the y axis, the horizontal polarization $|H\rangle$ is along the x axis, and the diagonal polarization $|D\rangle$ would be along the $y = x$ line. Thus, you would expect the diagonal polarization state to be proportional to a combination, or in quantum mechanics terms, a superposition of the vertical state and the horizontal state, represented by $|D\rangle \sim (|V\rangle + |H\rangle)$.

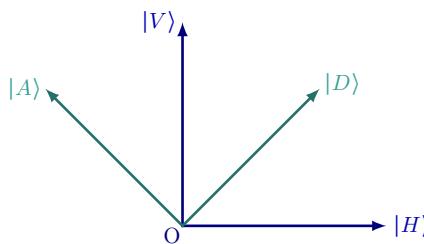


Figure 1.4: Relationship between H/V and D/A Polarizations

We have already assigned vectors to the vertical and horizontal polarizations, so the diagonal polarization should be proportional to $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Only we have a problem; this vector is not normalized, meaning if you take the inner product of this vector with itself, which gives you $1^2 + 1^2 = 2$, it is not 1. So, in order to make this vector normalized, we need to divide it by $\sqrt{2}$. This correction leads us to the actual representation of the diagonal state:

$$|D\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (1.9)$$

or in Dirac notation,

$$|D\rangle = \frac{1}{\sqrt{2}} (|V\rangle + |H\rangle). \quad (1.10)$$

Now, let's look at the anti-diagonal polarization state $|A\rangle$; it is 135° , along the $y = -x$ line. This polarization state, you can imagine it should be $|A\rangle \sim (|V\rangle - |H\rangle)$, which is $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Once again, this vector is not normalized. So, to normalize it, we need to divide it by $\sqrt{2}$, therefore

$$|A\rangle = \frac{1}{\sqrt{2}} (|V\rangle - |H\rangle) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \quad (1.11)$$

The D and A polarizations are also mutually exclusive, reflected as $|D\rangle$ and $|A\rangle$ being orthogonal, i.e., $\langle D|A\rangle = 0$.

■ **Example 1.1** Let's verify $\langle D|A\rangle = 0$. In matrix representation,

$$\langle D|A\rangle = \frac{1}{\sqrt{2}} [1 \ 1] \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0.$$

In Dirac notation,

$$\begin{aligned} \langle D|A\rangle &= \frac{1}{\sqrt{2}} (\langle V| + \langle H|) \frac{1}{\sqrt{2}} (|V\rangle - |H\rangle) \\ &= \frac{1}{2} (\langle V|V\rangle - \langle V|H\rangle + \langle H|V\rangle - \langle H|H\rangle) \\ &= \frac{1}{2}(1 - 0 + 0 - 1) = 0. \end{aligned}$$

Since $|D\rangle$ and $|A\rangle$ are orthonormal, they also form a basis for photon polarization states. A general polarization state can be represented in this basis as

$$|P\rangle = \gamma |D\rangle + \delta |A\rangle, \quad (1.12)$$

where γ and δ are complex numbers satisfying $|\gamma|^2 + |\delta|^2 = 1$.

In particular,

$$|V\rangle = \frac{1}{\sqrt{2}} (|D\rangle + |A\rangle), \quad (1.13a)$$

$$|H\rangle = \frac{1}{\sqrt{2}} (|D\rangle - |A\rangle). \quad (1.13b)$$

In the context of quantum computing, $|D\rangle$ and $|A\rangle$ correspond to the $|+\rangle$ and $|-\rangle$ states, which we will study later.

Exercise 1.2 Compare Equation 1.12 and Equation 1.6, and derive formulas for

- (a) γ and δ in terms of α and β

(b) α and β in terms of γ and δ

using both matrix representation and Dirac notation.

1.3.5 Circular Polarization States

So far, we have only examined linear polarizations. In linear polarization, the polarization vector stays fixed as the photon propagates. But in circular polarization, as depicted in Fig. 1.2, the polarization vector rotates as the photon propagates. With right circular polarization $|R\rangle$, the turning of the polarization vector is right-handed as we look in the direction of propagation. When the photon travels one wavelength, the polarization vector will turn by 360° clockwise.

Let's look at how we can construct a state vector for this quantum state. We'll utilize the xy coordinate system, as for photon polarizations, we have the luxury to do so. The difference between linear polarization $|D\rangle$ and circular polarization $|R\rangle$ is that for $|D\rangle$, the horizontal vector and the vertical vector are always in phase, but for $|R\rangle$, the state vector goes through the y axis first then turns to the x axis; or equivalently, the x ($|H\rangle$) lags the y ($|V\rangle$) component by a 90° phase.

So we need to introduce a phase factor. A 90° lagging phase corresponds to a phase factor $e^{-i\pi/2} = -i$. Therefore, we expect the state vector $|R\rangle$ to be

$$|R\rangle = \frac{1}{\sqrt{2}} (|V\rangle - i|H\rangle) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix}. \quad (1.14)$$

Again, the $\sqrt{2}$ factor comes from the normalization requirement.

Of course, we could also argue that the y component *leads* the x component by a 90° phase, and correspondingly $|R\rangle = \frac{1}{\sqrt{2}} (i|V\rangle + |H\rangle)$. This expression and Eq. 1.14 should be physically equivalent. Note that the two differ just by a global phase factor i :

$$(i|V\rangle + |H\rangle) = i(|V\rangle - i|H\rangle).$$

This leads to the following observation:

Implication 1.3: Global Phase

Global phase of a quantum state vector does not have physical significance. That is, $e^{i\phi}|\psi\rangle$ represents the same quantum state as $|\psi\rangle$, where ϕ is a real number representing a phase angle.

What about left-handed circular polarization? It is the opposite to right polarization, with the H component leading the V component by a 90° phase. So the state vector (up to a global phase) is

$$|L\rangle = \frac{1}{\sqrt{2}} (|V\rangle + i|H\rangle) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix}. \quad (1.15)$$

Exercise 1.3 Verify that $|L\rangle$ and $|R\rangle$ are normalized and orthogonal with each other, using both matrix representation and Dirac notation.

Similar to $|H\rangle$ and $|V\rangle$, and $|D\rangle$ and $|A\rangle$, $|L\rangle$ and $|R\rangle$ also form a basis for photon polarization states, because they are orthonormal. A general polarization state can be represented in this basis as

$$|P\rangle = \mu |L\rangle + \nu |R\rangle. \quad (1.16)$$

Exercise 1.4 Represent $|V\rangle$ and $|H\rangle$ in terms of $|L\rangle$ and $|R\rangle$.

1.3.6 General Polarization States

We have established that any polarization state, denoted by $|P\rangle$, can be represented in the form (see Eq. 1.6):

$$|P\rangle = \alpha |V\rangle + \beta |H\rangle. \quad (1.17)$$

Here, α and β are complex numbers subject to the normalization condition $|\alpha|^2 + |\beta|^2 = 1$. These coefficients can be parameterized by two real angles θ and ϕ as:

$$\alpha = \cos \theta, \quad \beta = \sin \theta e^{i\phi}. \quad (1.18)$$

The normalization condition is naturally satisfied due to $\cos^2 \theta + \sin^2 \theta = 1$, allowing for an arbitrary phase $e^{i\phi}$ between $|V\rangle$ and $|H\rangle$. Thus, $|P\rangle$ becomes:

$$|P\rangle = \cos \theta |V\rangle + \sin \theta e^{i\phi} |H\rangle. \quad (1.19)$$

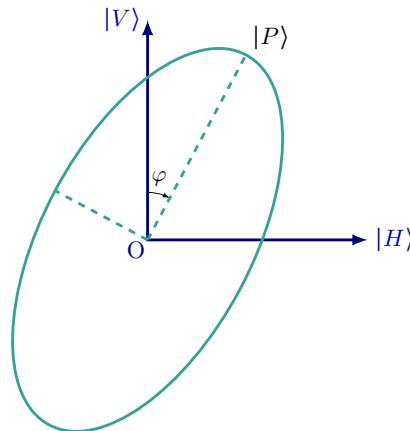


Figure 1.5: Elliptical Polarization

This general form accounts for elliptical polarization, as illustrated in Fig. 1.5. Rectilinear, diagonal, and circular polarizations emerge as special cases, as summarized in the table below.

Polarization State	θ	ϕ
$ V\rangle$	0	0
$ H\rangle$	$\pi/2$	0
$ D\rangle$	$\pi/4$	0
$ A\rangle$	$-\pi/4$	0
$ L\rangle$	$\pi/4$	$\pi/2$
$ R\rangle$	$\pi/4$	$-\pi/2$

Exercise 1.5 Validate that Eq. 1.19 reproduces the special cases listed above.

Exercise 1.6 Show that when $\phi = 0$ Eq. 1.19 represents a linear polarization state. What is its angle relative to $|V\rangle$?

We have previously seen that $|P\rangle$ can be alternatively expressed in different bases (Eqs. 1.12 and 1.16).

1.3.7 The Quantum State Postulate

In § 1.3.1 we have introduced the Quantum State Postulate in complex vector space, which is adequate for quantum computing in most cases. More generally, quantum states reside in a Hilbert space.

Postulate 1: Quantum State

The state of a quantum system is completely described by a normalized complex vector in a Hilbert space.

A Hilbert space (\mathcal{H}) is a vector space equipped with an inner product that is complete, meaning that it contains all its limit points. Its dimension is $N = 2^n$ for a system of n qubits, and can even be infinite. But for our purpose in quantum computing, we can consider a Hilbert space \mathcal{H} as a complex vector space, denoted as \mathbb{C}^N .

For a two-level system, the quantum state can be described by:

$$|\psi\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (1.20)$$

Here $\alpha, \beta \in \mathbb{C}$ satisfying $\langle\psi|\psi\rangle = |\psi|^2 = 1$, and the state space is \mathbb{C}^2 .

1.4 The Quantum Observable Postulate

As we delve into the intricate landscape of quantum mechanics, we encounter a key juncture where the State Postulate seamlessly gives way to the Measurement Postulate. Central to navigating this landscape are observables, fundamental quantities that lie at the core of understanding measurable properties within quantum systems. The Quantum Observable Postulate, our starting point, stands as the foundation of this exploration.

In the context of Quantum Computing, where harnessing quantum phenomena holds immense potential, measurement assumes paramount importance. Measurement serves as the gateway through which we extract information from quantum systems, shaping the very essence of Quantum Computing.

Our journey begins with the Observable Postulate, followed by the subsequent unveiling of the Measurement Postulate. Each step reveals distinct facets of the quantum realm. Through a series of subsections, we unravel the transition from classical macroscopic quantities to the enigmatic world of quantum observables. Along this path, we delve into illustrative examples, unravel the intricacies of mathematical representation, and unveil the profound interplay between observables and quantum states. This section, serving as both a guide and an exploration, establishes the groundwork for a deeper voyage into the boundless potential of Quantum Computing.

1.4.1 Macroscopic Quantities to Quantum Observables

The principle of quantum states has been introduced through illustrative examples involving photon polarization. According to this principle, the state of a quantum system finds complete description within a normalized state vector. It is worth noting the term ‘completely,’ which we’ve hitherto left unexplored. Additionally, we’ve established that the components of these state vectors are generally complex numbers.

One might inquire: how do physical attributes like light intensity, energy, position, and velocity fit into this framework? These attributes are categorized as quantum observables (also referred to as measurables), and they constitute the focus of our forthcoming exploration.

In quantum mechanics, observables are represented by Hermitian operators, symbolized as M in the subsequent discussion. In this context, M can signify properties such as position or light intensity. These operators take the form of matrices, possessing eigenvalues and eigenvectors that relate to the measurable outcomes attainable from the quantum state system.

Postulate 2: Quantum Observables

Any physically measurable quantity is represented by a Hermitian operator (matrix) M .

The eigenvalues of M , denoted as $\{\lambda_i\}$, represent the possible outcomes when measuring M .

For each eigenvalue λ_i , the corresponding eigenvector $|\lambda_i\rangle$ is the state that yields that particular measurement outcome with certainty.

The Observable Postulate is illustrated in Fig. 1.6. This postulate states that any physically measurable quantity in quantum mechanics can be associated with a Hermitian operator known as an observable, represented here by M . When this operator acts on its eigenstate $|\lambda_i\rangle$, the outcome is the state itself scaled by a corresponding eigenvalue λ_i . For a d -dimensional system (i.e., the operator M is a $d \times d$ matrix), there are d such pairs of eigenvalues and eigenvectors, and the index i runs from 1 to d , accounting for each dimension of the system. Eigenvalues

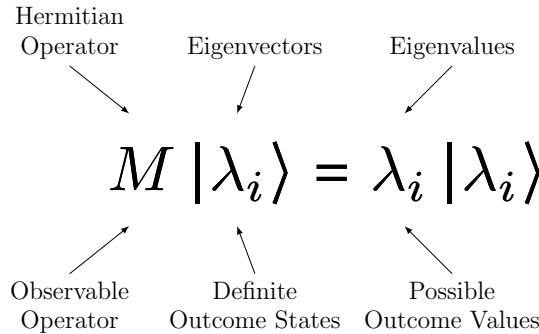


Figure 1.6: Illustration of the Quantum Observable Postulate

may be degenerate, meaning different eigenvectors can correspond to the same eigenvalue. For further details on eigenvectors and eigenvalues in the context of quantum observables, see Appendix C and Ref [9].

This concept may appear quite mathematical, but it will become clearer with examples that we will explore next. In quantum computing, this postulate informs how we can extract information from qubits. The readout is essentially a measurement, and thus the postulate helps in understanding how algorithms will behave when translated into physical operations.

Since quantum observables are Hermitian operators (matrices), let's first review the properties of those in order to understand this postulate:

- Definition of Hermitian operator: $M^\dagger = M$.
- Hermitian operators have real eigenvalues, ensuring that measurements yield real-numbered outcomes.
- The eigenstates associated with distinct eigenvalues of a Hermitian operator are guaranteed to be orthogonal. (A mathematical nuance: while the eigenvalues are unique, the eigenvectors corresponding to them are not. Specifically, non-degenerate eigenvectors can differ by an overall phase factor, which has no observable physical consequence.)
- Hermitian operators possess a complete set of eigenstates, allowing any quantum state to be expanded as a linear combination of these eigenstates.

Exercise 1.7 For any complex matrix A , prove that AA^\dagger is a Hermitian matrix.

Exercise 1.8 Find the eigenvalues and eigenvectors for

$$\begin{bmatrix} a - \alpha & b + i\beta \\ b - i\beta & a + \alpha \end{bmatrix},$$

where $a, b, \alpha, \beta \in \mathbb{R}$.

1.4.2 Observables for Rectilinear Polarizations

Let's examine linear polarizations again. According to the observable principle, the reason we can measure these states with definite answers (pass or blocked) means that there is an observable for each case. That observable is a particular Hermitian matrix, and the eigenvectors of this matrix should be the vectors corresponding to those definite states. We will demonstrate how this works.

Let's start with the vertical polarization $|V\rangle$, and we use the vertical observable M_V to measure it:

$$M_V = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \quad (1.21)$$

We know when we use a vertical polarizer to measure a photon with vertical polarization, we get a definite yes answer. If we assign a value 1 to yes (pass) and 0 to no (blocked) to the measurement outcome, we expect $|V\rangle$ to be an eigenvector of M_V with an eigenvalue 1:

$$M_V |V\rangle = 1 |V\rangle = |V\rangle, \quad (1.22)$$

which you can easily verify. Similarly, when we use a vertical polarizer to measure a photon with horizontal polarization, we get a definite no answer. There is no maybe answer. We expect $|H\rangle$ to be an eigenvector of M_V with an eigenvalue 0:

$$M_V |H\rangle = 0 |H\rangle = 0. \quad (1.23)$$

This demonstrates an important concept derived from the Quantum Observable principle:

Implication 2.1: Quantization of Measurement Outcomes

Unlike classical systems where observables can take continuous ranges of values, quantum observables often have discrete eigenvalues.

The same concept also underlies why atoms have discrete energy levels, and that qubits exist as two-level systems. It is explained by the observable principle because only possible outcome of a quantum measurement is one of the eigenvalues of the corresponding observable operator. These eigenvalues are often discrete numbers.

Similarly, if we use a horizontal polarizer to measure vertical and horizontal polarizations, we are using an observable M_H (which is complimentary to M_V):

$$M_H = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad (1.24)$$

$$M_H |V\rangle = 0, \quad (1.25a)$$

$$M_H |H\rangle = |H\rangle. \quad (1.25b)$$

1.4.3 Observables for Diagonal Polarizations

Now, let's look at the diagonal polarizations $|D\rangle$ and $|A\rangle$. They are also mutually exclusive quantum states, and the two vectors are orthonormal. We will measure them using a diagonal polarizer, represented by an observable operator M_D . This round, instead of verifying the observable matrix, let's derive it, which involves essential skills for anyone working on quantum computing. Let

$$M_D = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad (1.26)$$

and we require

$$M_D |D\rangle = |D\rangle, \text{ i.e., } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (1.27a)$$

$$M_D |A\rangle = 0, \text{ i.e., } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0. \quad (1.27b)$$

Through some algebra, we can determine that $a = b = c = d = \frac{1}{2}$, so:

$$M_D = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \quad (1.28)$$

The preceding demonstration of M_V and M_D leads to another implication of the Observable Postulate:

Implication 2.2: Measurement Basis

The choice of observable determines the basis in which the measurement outcome is to be expressed.

Exercise 1.9 Derive the observable operator M_A for measuring $|D\rangle$ with outcome 0 and $|A\rangle$ with outcome 1, similar to how M_D is derived in Eq. 1.27.

1.4.4 Observables for Circular Polarizations

The circular polarizations $|L\rangle$ and $|R\rangle$ are also mutually exclusive quantum states. There should be an observable operator M_L which has $|L\rangle$ as eigenvector with eigenvalue 1, and $|R\rangle$ with eigenvalue 0; and vice versa for the complimentary observable M_R .

This round, we will demonstrate a more general method to derive the matrices of M_L and M_R . According to the spectral decomposition theorem, a Hermitian matrix can be expressed as the sum of its eigenvalues times the outer products of the eigenvectors:

$$M = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|. \quad (1.29)$$

Therefore,

$$M_L = 1 |L\rangle\langle L| + 0 |R\rangle\langle R| \quad (1.30a)$$

$$= \frac{1}{\sqrt{2}} (|V\rangle + i|H\rangle) \frac{1}{\sqrt{2}} (\langle V| - i\langle H|) \quad (1.30b)$$

$$= \frac{1}{2} (|V\rangle\langle V| - i|V\rangle\langle H| + i|H\rangle\langle V| + |H\rangle\langle H|), \quad (1.30c)$$

or, in matrix form,

$$M_L = \frac{1}{2} \begin{matrix} & \langle V| & \langle H| \\ |V\rangle & \left[\begin{array}{cc} 1 & -i \\ i & 1 \end{array} \right] \\ |H\rangle & \end{matrix}. \quad (1.31)$$

The above derivation of M_L is a key skill for anyone learning quantum computing. Through M_L we have also demonstrated that even though an observable operator may contain complex numbers, its associated eigenvalues are real numbers, owing to a property of Hermitian matrices.

Exercise 1.10 Using both the matrix representation and Dirac notation, verify that

$$\begin{aligned} M_L |L\rangle &= |L\rangle, \\ M_L |R\rangle &= 0. \end{aligned}$$

Exercise 1.11 Derive the observable operator M_R for measuring $|L\rangle$ with outcome 0 and $|R\rangle$ with outcome 1, similar to how M_L is derived in Eq. 1.30.

1.4.5 Basis Equivalence and Dirac Notation

Quantum states and observables are not tied to a specific basis; rather, their vector or matrix representation changes when expressed in different bases. According to the Observable Postulate, each observable's eigenvectors serve as a basis for quantum states. These bases, formed by the eigenvectors of Hermitian operators, are mathematically interchangeable for describing quantum systems.

Dirac's bra-ket notation offers a consistent, basis-agnostic framework for representing these quantum entities. For a succinct review of Dirac notation in the context of linear algebra, refer to Appendix C.

Basis-Independence in Dirac Bra-Ket Notation

The Dirac bra-ket notation abstractly and universally represents quantum states and operators, highlighting their inherent basis-independence.

We previously examined photon polarization states and their associated observables. Table 1.2 consolidates their vector and matrix representations in both rectilinear and diagonal bases.

As a demonstration of the power of the Dirac notation, let's derive the matrix of M_L in the $\{|D\rangle, |A\rangle\}$ basis. We treat the equations in Table 1.2 as basis-independent relations, and work through the algebra to obtain M_L in terms of $|D\rangle$ and $|A\rangle$:

Relations			
	$\{ V\rangle, H\rangle\}$	$\{ D\rangle, A\rangle\}$	
$ D\rangle = \frac{1}{\sqrt{2}}(V\rangle + H\rangle)$		$ A\rangle = \frac{1}{\sqrt{2}}(V\rangle - H\rangle)$	
$ L\rangle = \frac{1}{\sqrt{2}}(V\rangle + i H\rangle)$		$ R\rangle = \frac{1}{\sqrt{2}}(V\rangle - i H\rangle)$	
$M_V = V\rangle\langle V $		$M_D = D\rangle\langle D $	$M_L = L\rangle\langle L $

Quantity	$\{ V\rangle, H\rangle\}$ Basis	$\{ D\rangle, A\rangle\}$ Basis
$ V\rangle$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$
$ H\rangle$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$
$ D\rangle$	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$
$ A\rangle$	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$
$ L\rangle$	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ i \end{bmatrix}$	$\frac{1}{2}\begin{bmatrix} 1+i \\ 1-i \end{bmatrix}$
$ R\rangle$	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -i \end{bmatrix}$	$\frac{1}{2}\begin{bmatrix} 1-i \\ 1+i \end{bmatrix}$
M_V	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	$\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
M_D	$\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
M_L	$\frac{1}{2}\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}$	$\frac{1}{2}\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$

Table 1.2: Summary of Photon Polarization States and Observables

$$M_L = |L\rangle\langle L| \quad (1.32a)$$

$$= \frac{1}{\sqrt{2}} (|V\rangle + i|H\rangle) \frac{1}{\sqrt{2}} (\langle V| - i\langle H|) \quad (1.32b)$$

$$= \frac{1}{2} ((|D\rangle + |A\rangle) + i(|D\rangle - |A\rangle)) \frac{1}{2} ((\langle D| + \langle A|) - i(\langle D| - \langle A|)) \quad (1.32c)$$

$$= \frac{1}{2} (|D\rangle\langle D| + i|D\rangle\langle A| - i|A\rangle\langle D| + |A\rangle\langle A|) \quad (1.32d)$$

$$= \frac{1}{2} \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}. \quad (1.32e)$$

Exercise 1.12 In our standard formulation, matrices M_V , M_D , and M_L have eigenvalues of 1 (pass) for the first eigenvector and 0 (blocked) for the second. Consider an alternative assignment where the second eigenvector has an eigenvalue of -1 instead of 0. Derive the matrices M_V , M_D , and M_L under this new scheme, and represent them in the $|V\rangle, |H\rangle$ basis. You will discover the Pauli matrices.



While quantum computing commonly employs orthonormal basis states tied to specific observables, it's worth noting that mathematically, basis states need only to be linearly independent and complete; orthonormality is not a requirement.

1.5 The Quantum Measurement Postulate

The Observable Postulate and the Measurement Postulate constitute fundamental tenets within the realm of quantum mechanics. These postulates delineate the behavior of quantum systems and the intricate process of measurement. Specifically, the Observable Postulate stipulates that every observable quantity, such as position and light intensity, is symbolized by a Hermitian operator in quantum mechanics. On the other hand, the Measurement Postulate details the methodology of measuring an observable quantity within a quantum system.

In the ensuing discussion, we will delve into the intricacies of the Measurement Postulate, dissecting it into two key components: measurement probability and state collapse. These principles form the bedrock of quantum mechanics, providing insight into the interdependencies between measurements, states, and probabilities. We will elucidate these concepts using the context of photon polarization, employing it as a visual aid to comprehend the embodiment of these principles within quantum systems.



Quantum measurement is a complex subject that continues to evolve with advancements in quantum computing and quantum information science. The standard quantum measurement postulates presented in this section serve as foundational concepts. These will be extended in § 3.4 to encompass a general measurement framework, including more complex scenarios such as Positive Operator-Valued Measures (POVMs), thereby offering increased flexibility and broader applicability.

1.5.1 Measurement Probability

The Observable Postulate informs us that if a quantum state takes the form of an

eigenstate $|\lambda_i\rangle$ of an observable M , the measurement outcome will unequivocally yield λ_i . In essence, distinguishing eigenstates of an observable remains unambiguous.

However, when we endeavor to measure a general quantum state—such as a superposition state $|\psi\rangle$ as presented in Eq. 1.17—that does not conform to an eigenstate of M , multiple outcomes come into play, inherently characterizing quantum measurements as probabilistic phenomena.

Postulate 3a: Measurement Probability

The outcome of measuring an observable M on a quantum state $|\psi\rangle$ corresponds to one of its eigenvalues. The probability associated with each eigenvalue λ_i is given by $|\langle\lambda_i|\psi\rangle|^2$ (the Born rule).

This postulate offers a multi-faceted understanding of quantum measurement, leading to several significant implications, as detailed below:

Implication 3.1: Non-deterministic Outcomes

Quantum measurements inherently entail probabilistic outcomes—a distinct contrast to classical physics where measurements are deterministic provided knowledge of both initial conditions and governing equations.

Implication 3.2: Quantum State and Measurement Probabilities

The Born rule establishes the mathematical framework that links quantum states to the probabilities of measurement outcomes.

Implication 3.3: Total Probability

Quantum state vectors maintain normalization across all bases, aligning with the fundamental requirement that probabilities of mutually exclusive events must collectively sum to 1.

Photon Polarization Experiments

The experiments shown in Fig. 1.7 exemplify the Born rule. Photons are prepared in the vertical polarization $|\psi\rangle = |V\rangle$. In part (a), they are measured with a vertical polarizer. The passing probability is $|\langle V|\psi\rangle|^2 = |\langle V|V\rangle|^2 = 1$, thus 100% of the photons pass through.

In part (b), where the photons are measured with a horizontal polarizer, the passing probability becomes $|\langle H|\psi\rangle|^2 = |\langle H|V\rangle|^2 = 0$, and consequently, no photons pass through.

In part (c), when the photons are measured with a diagonal polarizer, the passing probability is $|\langle D|\psi\rangle|^2 = |\langle D|V\rangle|^2 = 0.5$. As a result, 50% of photons pass through while the rest are blocked. This is reflected in the figure by the doubled spacing between the photons.

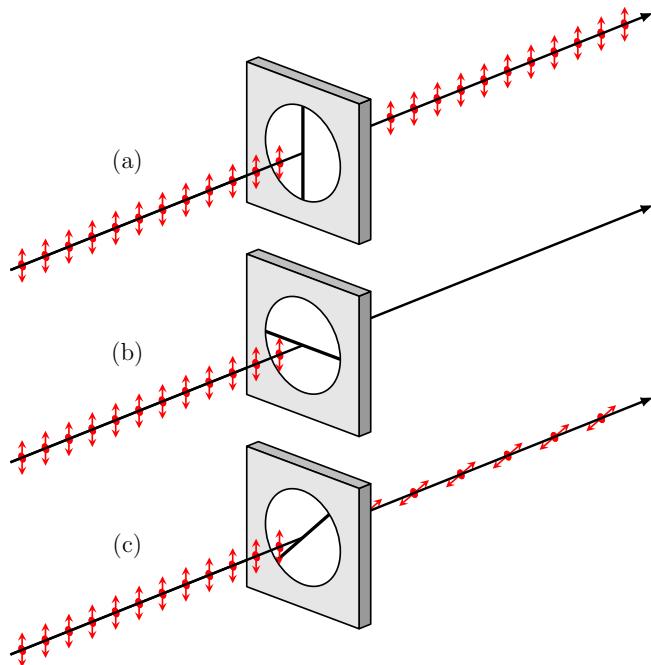


Figure 1.7: Measuring Photons Through a Polarizer

Exercise 1.13 What is the passing probability of vertically polarized photons $|\psi\rangle = |V\rangle$ when measured with a device that passes left circular polarization M_L ?

1.5.2 State Collapse

In classical physics, measurements can often be conducted in such a way that the system remains undisturbed. The deterministic framework of classical physics enables these non-invasive measurements. However, quantum mechanics is fundamentally different. Owing to the quantization of energy and other properties, along with the limitations imposed by the Heisenberg Uncertainty Principle, any measurement in a quantum system inevitably involves a disturbance to that system. For example, measuring an electron's position necessarily affects its momentum. This departure from classical intuition is one of the defining traits of quantum mechanics.

This divergence between classical and quantum paradigms leads us to an essential feature of quantum theory: the postulate of measurement state collapse. In quantum systems, a measurement forces a transition in the state of the system. Specifically, according to Postulate 3b, when an observable M is measured on a quantum state $|\psi\rangle$, the state collapses to one of the eigenstates $|\lambda_i\rangle$ corresponding to the measured eigenvalue λ_i . This state collapse is immediate and irreversible.

Postulate 3b: Measurement State Collapse

After the measurement, the quantum state collapses to the eigenstate associated with the measured eigenvalue.

The immediate implication of this postulate is that measurements in quantum mechanics are fundamentally non-reversible processes.

Implication 3.4: Irreversible Effect

Measurements permanently alter the quantum state, forcing it into one of the eigenstates of the measured observable. This non-reversibility is non-classical and affects the system's subsequent dynamics.

In other words, the act of measurement irrevocably changes the quantum state. Once the system is in an eigenstate of the observable M , subsequent measurements with the same observable will consistently yield the same result.

Implication 3.5: Subsequent Measurements

Once a measurement has been made, the system's state collapses to a definite eigenstate. Therefore, any repeated measurements with the same observable will yield identical results.

The Copenhagen Interpretation and von Neumann Framework

The measurement state collapse postulate aligns with the Copenhagen interpretation of quantum mechanics, primarily attributed to Niels Bohr and Werner Heisenberg. John von Neumann later provided a mathematically rigorous version of this perspective, delineating deterministic unitary evolution from non-deterministic measurement-induced collapse. While several interpretations of quantum mechanics exist, in this text, we adopt the von Neumann framework. Further discussions on various interpretations can be found in § 9.2.

Examples

Repeated Measurements with Photon Polarization

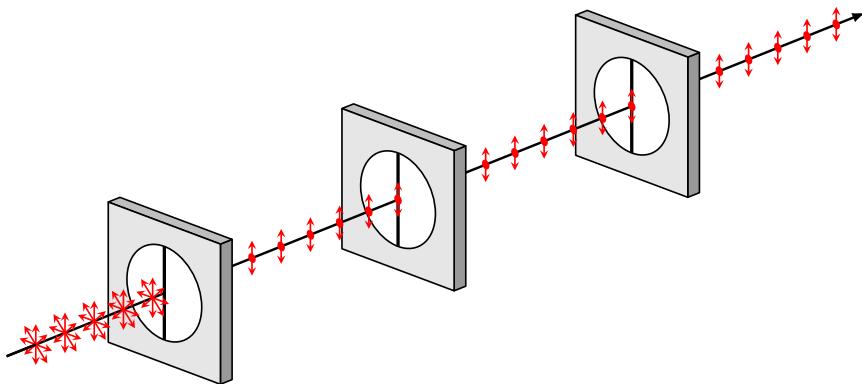


Figure 1.8: Repeated Identical Measurements

Figure 1.8 serves as a concrete example of this principle. Initially, the photons are unpolarized, implying a mixture of different polarization states. (We will discuss such mixed quantum states and their representation by density matrices in a later section.) When the photons pass through the first vertical polarizer, their states collapse to vertical polarization. Subsequent measurements with vertical polarizers will then result in a 100% transmission rate, without blocking or altering the photon states.

Cross-Polarizer Experiments

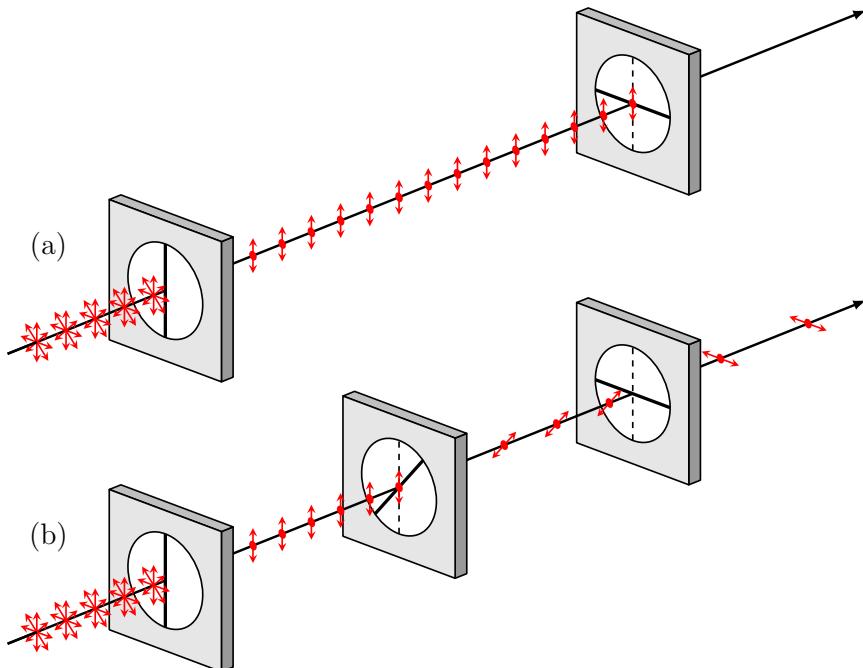


Figure 1.9: Measuring Photons Through Crossed Polarizers

Armed with the Measurement Postulate, we can elucidate the perplexing cross-polarizer experiment discussed in § 1.2.3. Fig. 1.9 shows the quantum rendition of this experiment, where light is treated as a photon stream.

In part (a), we introduce unpolarized photons to a vertical polarizer. Approximately 50% of the photons pass, consistent with a representation of unpolarized photons as being 50% in $|V\rangle$ and 50% in $|H\rangle$. Photons that do pass are forced into the vertical polarization state $|V\rangle$. Subsequent measurement using a horizontal polarizer yields a zero probability of passage, $|\langle H|V \rangle|^2 = 0$, aligning with experimental findings.

In part (b), a diagonal polarizer is inserted between the vertical and horizontal polarizers. The probability of photons in state $|V\rangle$ passing this diagonal polarizer is $|\langle V|D \rangle|^2 = 50\%$. The photons that pass are then in state $|D\rangle$. Measuring these photons with a horizontal polarizer gives a passage probability of 50%, resulting in an overall probability of $50\% \times 50\% \times 50\% = 12.5\%$, which matches the experimental observations.

The crux of this experiment lies in the Measurement Postulate-induced state collapse to $|D\rangle$ upon passing the diagonal polarizer. Absent this quantum mechanical feature, the photons would not navigate through all three polarizers in part (b), especially considering they are entirely blocked by just two polarizers in part (a).

1.5.3 *Statistical Average

In quantum mechanics, physical observables are represented by Hermitian operators M . The macroscopic measurements we make yield the statistical average of these observables, calculated in accordance with the postulates discussed earlier.

Consider the eigenvectors $|\lambda_i\rangle$ of M . They form an orthonormal basis, allowing us to express a general quantum state as

$$|\psi\rangle = \sum_i c_i |\lambda_i\rangle, \quad (1.33)$$

where the coefficients c_i are complex numbers satisfying $\sum_i |c_i|^2 = 1$.

Upon measuring a superposition state using the operator M , the result is probabilistic. The possible outcomes are the eigenvalues λ_i , each occurring with a probability $P_i = |\langle\lambda_i|\psi\rangle|^2$. The statistical average of these outcomes can be calculated as:

$$\langle M \rangle = \sum_i P_i \lambda_i = \sum_i |\langle\lambda_i|\psi\rangle|^2 \lambda_i. \quad (1.34)$$

Interestingly, this can be simplified to a basis-independent form:

Implication 3.6: Statistical Average

The statistical average (expected value, or expectation value) for measuring an observable M on a quantum state $|\psi\rangle$ is given by:

$$\langle M \rangle = \langle \psi | M | \psi \rangle. \quad (1.35)$$

To build intuition for this expression, consider the case where $|\lambda_0\rangle = |0\rangle$, $|\lambda_1\rangle = |1\rangle$, and $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$.

$$\begin{aligned} \langle M \rangle &= |\alpha|^2 \lambda_0 + |\beta|^2 \lambda_1 \\ &= \langle \psi | 0 \rangle \langle 0 | \psi \rangle \lambda_0 + \langle \psi | 1 \rangle \langle 1 | \psi \rangle \lambda_1 \\ &= \langle \psi | (\lambda_0 |0\rangle\langle 0| + \lambda_1 |1\rangle\langle 1|) | \psi \rangle \\ &= \langle \psi | M | \psi \rangle, \end{aligned}$$

where we have used the spectral decomposition property of Hermitian matrix M :

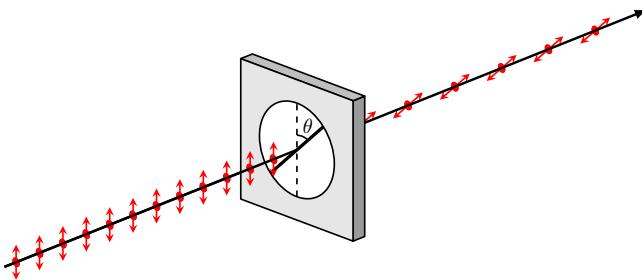
$$M = \sum_i \lambda_i |\lambda_i\rangle \langle \lambda_i|. \quad (1.36)$$

The general proof is similar to our simple example:

Proof.

$$\begin{aligned}
\langle M \rangle &= \sum_i |\langle \lambda_i | \psi \rangle|^2 \lambda_i \\
&= \sum_i \langle \psi | \lambda_i \rangle \langle \lambda_i | \psi \rangle \lambda_i \\
&= \sum_i \langle \psi | (\lambda_i |\lambda_i\rangle \langle \lambda_i|) |\psi \rangle \\
&= \langle \psi | \left(\sum_i \lambda_i |\lambda_i\rangle \langle \lambda_i| \right) |\psi \rangle \\
&= \langle \psi | M | \psi \rangle.
\end{aligned}$$

□

Example with Photon Polarization: Malus' Law**Figure 1.10:** Measuring Photons: Malus' Law

To illustrate Eq. 1.35, we consider a scenario involving vertically polarized photons measured with a polarizer oriented at an angle θ , as depicted in Fig. 1.10. The state of the photon after passing through the polarizer can be described by

$$|\theta\rangle = \cos \theta |V\rangle + \sin \theta |H\rangle = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}. \quad (1.37)$$

The associated observable operator for the polarizer is given by

$$M_\theta = |\theta\rangle \langle \theta| = \begin{bmatrix} \cos^2 \theta & \cos \theta \sin \theta \\ \cos \theta \sin \theta & \sin^2 \theta \end{bmatrix}. \quad (1.38)$$

To find the statistical average of M_θ , we apply Eq. 1.34:

$$\langle M \rangle = |\langle \theta | V \rangle|^2 \cdot 1 = \cos^2 \theta. \quad (1.39)$$

Alternatively, using Eq. 1.35, we obtain

$$\langle M \rangle = \langle V | M_\theta | V \rangle = \cos^2 \theta. \quad (1.40)$$

Both methods yield the statistical average as $\cos^2 \theta$, which confirms Malus' Law: the output-to-input intensity ratio is $\cos^2 \theta$. Since light intensity is proportional to photon flux, this result provides a quantum-mechanical rationale for Malus' law.

One may ask why we should prove Malus' law using quantum mechanics when it can be explained using classical electromagnetic theory. There are several compelling reasons. Firstly, using photons as an example helps develop a foundational understanding of quantum mechanics. Secondly, the quantum theory of photons is essential for exploring advanced quantum phenomena such as photon entanglement. Lastly, this example serves to illustrate how classical physics emerges as a statistical limit of quantum physics.

Exercise 1.14 Suppose we measure circularly polarized photons $|L\rangle$ with the polarizer M_θ . Calculate the statistical average for this case.

Exercise 1.15 Suppose we measure $|V\rangle$ photons using a specialized device D rather than a polarizer. This device yields a $+1$ for photons polarized at an angle θ and -1 for the orthogonal polarization at $\theta + \pi/2$. Calculate $\langle D \rangle$.

1.5.4 Quantum Interference

Quantum mechanics fundamentally challenges our classical understanding of waves and particles with its concept of wave-particle duality. This is beautifully encapsulated by the phenomenon of interference. Unlike classical interference, which involves the superposition of wave amplitudes, quantum interference incorporates the quantum State Postulate, which permits particles like photons to exist in superposition states, and the Observables and Measurements Postulates, which govern the measurement of these superpositions.

Quantum Interference

Quantum interference is the phenomenon where particles like photons exist in superposition states and interfere with themselves upon measurement.

The simplest manifestation of quantum interference is the two-path interference. This phenomenon occurs when particles like photons or electrons have the potential to traverse two or more paths simultaneously and interfere with themselves. This mechanism serves as a cornerstone for studying quantum interference, entanglement, and the implementation of qubits.

1 Mach-Zehnder Interferometer

The Mach-Zehnder interferometer, depicted in Fig. 1.11, is an exemplary experimental setup to observe photon two-path interference. Upon entering the interferometer, the photon encounters a beam splitter (labeled BS1), which is essentially a half mirror – with 50% transmission and 50% reflection. This device effectively splits the photon into two paths: the transmitted path A and the reflected path B .



It is important to note a single photon can travel along two paths simultaneously, a phenomenon related to particle-wave duality. It is not that some photons travel in path A while others travel in path B . This has been extensively verified using “quantum eraser experiments” [61].

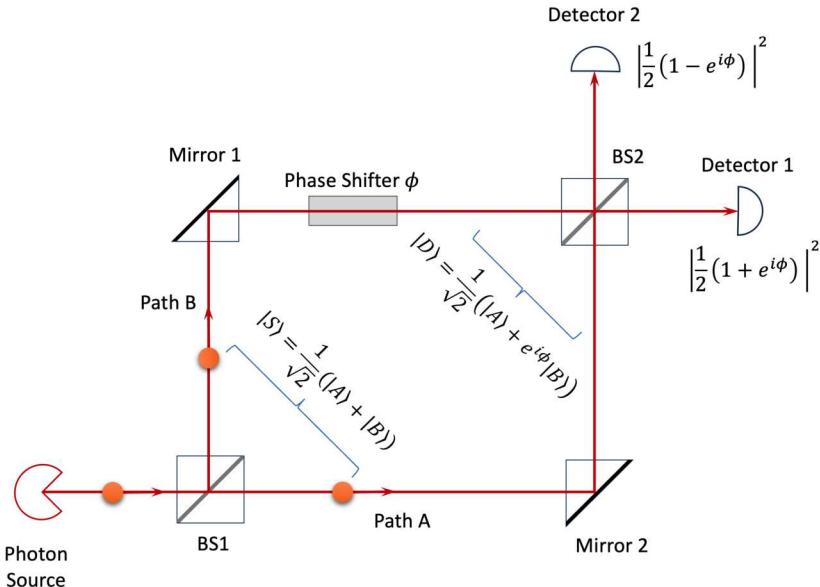


Figure 1.11: Mach-Zehnder Interferometer

The post-splitting superposition state is given by

$$|S\rangle = \frac{1}{\sqrt{2}}(|A\rangle + |B\rangle), \quad (1.41)$$

where $|A\rangle$ and $|B\rangle$ represent the photon states along paths A and B , respectively.

Following the first beam splitter, the photon is in a superposition state and may accrue a phase shift ϕ along path B relative to path A . This phase shift modifies the state to

$$|D\rangle = \frac{1}{\sqrt{2}}(|A\rangle + e^{i\phi}|B\rangle). \quad (1.42)$$

After traversing paths A and B , the photon encounters the second beam splitter (BS2). BS2 mixes the photon state from paths A and B and sends it to Detector 1 and Detector 2. The photon detectors, coupled with BS2, effectively measure the photon state in the basis $\frac{1}{\sqrt{2}}(|A\rangle + |B\rangle) \equiv |S\rangle$ and $\frac{1}{\sqrt{2}}(|A\rangle - |B\rangle)$, commonly denoted as $|+\rangle$ and $|-\rangle$ in quantum computing. (The beam splitter itself is functionally equivalent to a Hadamard gate, a topic to be elaborated on later.) Specifically, the observable operator associated with Detector 1 in the figure is $M_S = |S\rangle\langle S|$, which leads to the following probability for detecting the photon in state $|S\rangle$:

$$P = |\langle S|D\rangle|^2 = \left| \frac{1}{2}(1 + e^{i\phi}) \right|^2 = \cos^2 \frac{\phi}{2}. \quad (1.43)$$

Notice that P shows a sinusoidal dependence on ϕ . If $\phi = \pi$, the detection probability drops to zero, illustrating the principle of destructive interference.

Exercise 1.16 Using the quantum formulation of two-path interference, calculate the interference pattern in a Mach-Zehnder interferometer if (a) only the first beam splitter, and (b) both beam splitters, have a split ratio of 40-60 instead of 50-50.

2 Double-Slit Experiment

A related demonstration of photon two-path interference is the double-slit experiment, as depicted in Figure 1.12, where a single photon is fired at a barrier with two slits. Even though it's a single photon, it acts as if it has gone through both slits simultaneously, creating an interference pattern on a detector screen beyond the barrier.

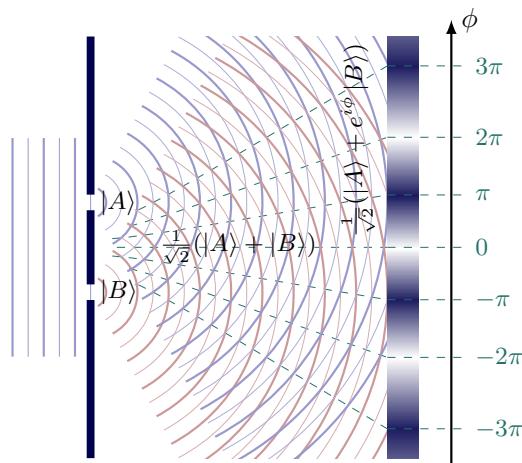


Figure 1.12: Double Slit Interference Experiment

Exercise 1.17 *Investigate the delayed-choice quantum eraser experiments, which are related to the double-slit experiment but delve into more captivating quantum effects associated with photon interference. Recommended sources include, but are not limited to, https://en.wikipedia.org/wiki/Quantum_eraser_experiment and https://en.wikipedia.org/wiki/Delayed-choice_quantum_eraser.

1.6 The Uncertainty Principle

The Measurement Postulate in quantum mechanics leads to several profound implications, one of which is the Uncertainty Principle. Though historically known as the Heisenberg Uncertainty Principle (HUP), it can be interpreted as a direct corollary of the Measurement Postulate.

The Uncertainty Principle has two forms. The first form distinguishes between compatible and incompatible observables. The second form presents the principle as a mathematical inequality. We will focus on the first version here, given its relevance to quantum computing.

1.6.1 Compatible versus Incompatible Observables

When performing a quantum measurement, the measured value is always an eigenvalue of the associated observable operator. Subsequently, the quantum state collapses to the corresponding eigenstate. Eigenstates of an operator represent states where measurement outcomes are certain, occurring with probability 1.

Two observables can either be compatible or incompatible, indicating whether both can be measured with 100% accuracy. This is often referred to as simultaneous measurements, although not necessarily at the same moment in time. Incompatibility between observables implies that measurements of one observable disturb those of the other. The Uncertainty Principle encapsulates this condition as follows:

The Uncertainty Principle

If operators A and B , representing two observables, do not commute, i.e., $[A, B] \neq 0$, then these observables cannot be measured simultaneously with perfect accuracy (i.e., with probability 1). In other words, they are incompatible.

Proof. Assume, for the sake of contradiction, that A and B can be measured simultaneously with perfect accuracy. Then, there must exist a complete set of simultaneous eigenstates $\{|\phi_i\rangle\}$, satisfying:

$$A|\phi_i\rangle = a_i|\phi_i\rangle, \quad B|\phi_i\rangle = b_i|\phi_i\rangle, \quad (1.44)$$

where a_i and b_i are the eigenvalues of A and B respectively.

If A and B possess simultaneous eigenstates $|\phi_i\rangle$, we have:

$$AB|\phi_i\rangle - BA|\phi_i\rangle = a_i b_i |\phi_i\rangle - b_i a_i |\phi_i\rangle = 0. \quad (1.45)$$

This implies that the commutator $[A, B] = AB - BA = 0$, contradicting our initial assumption $[A, B] \neq 0$.

Therefore, if A and B do not commute, they cannot have simultaneous eigenstates and thus cannot be measured simultaneously with perfect accuracy. \square

Implications of the Uncertainty Principle

Obersrvables A and B can be measured simultaneously

$$\Leftrightarrow [A, B] = 0$$

$$\Leftrightarrow A \text{ and } B \text{ share eigenvectors } \{|\phi_i\rangle\}$$

$$\Leftrightarrow A = \sum_i a_i |\phi_i\rangle\langle\phi_i|, \quad B = \sum_i b_i |\phi_i\rangle\langle\phi_i|.$$

Examples with Photon Polarization

To better understand the Uncertainty Principle in action, we can consider photon polarization as an illustrative example.

Vertical and Horizontal Polarization

Suppose we aim to measure the vertical polarization $|V\rangle$ and the horizontal polarization $|H\rangle$. Do these observables commute? Indeed, they do:

$$A \equiv M_V = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$B \equiv M_H = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

$$[A, B] = AB - BA = 0.$$

This means we can measure both $|V\rangle$ and $|H\rangle$ with 100% certainty.

Vertical and Diagonal Polarization

Now, consider measuring the vertical polarization $|V\rangle$ and the diagonal polarization $|D\rangle$:

$$A \equiv M_V = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$B \equiv M_D = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

$$[A, B] = AB - BA \neq 0.$$

Since $[A, B] \neq 0$, the observables for $|V\rangle$ and $|D\rangle$ are incompatible. In practice, if we measure $|V\rangle$ with 100% certainty, subsequent measurements for $|D\rangle$ will show both pass and block with 50% probability.

Applications in Quantum Communication: BB84 Protocol

In the context of the BB84 Quantum Key Distribution (QKD) protocol, consider a string of photons with varying polarizations.

- Case 1: Vertical and Horizontal - A sequence like $\{H, V, V, V, H, V, H, \dots\}$ can be measured with either a vertical or horizontal polarizer, confirming the polarization of each photon with 100% certainty.
- Case 2: Diagonal - For a sequence $\{D, A, D, A, A, A, D, \dots\}$, a 45° or 135° diagonal polarizer can measure the polarization states, since the states are orthogonal.
- Case 3: Mixture - If the sequence mixes vertical, horizontal, and diagonal polarizations, like $\{H, V, D, V, H, A, D, H, \dots\}$, perfect measurement is not possible. This stems from the incompatibility of the observable operators for $|V\rangle$ and $|D\rangle$, as shown above.

Exercise 1.18 Show that the measurements of left circular polarization $|L\rangle$ and the diagonal polarization $|D\rangle$ are incompatible by demonstrating that their observable operators do not commute. Explain what happens when photons in state $|D\rangle$ are measured for circular polarization and diagonal polarization.

1.6.2 *The Uncertainty Inequality

The Uncertainty Principle can also be formally expressed in terms of standard deviations of measurements for the operators A and B . This formulation is particularly

historical, often presented in the context of position and momentum or energy and time.

$$\Delta A \cdot \Delta B \geq \frac{1}{2} | \langle [A, B] \rangle |, \quad (1.46)$$

where $\Delta A = \sqrt{\langle A^2 \rangle - \langle A \rangle^2}$ and similarly for ΔB . Here, $[A, B] = AB - BA$ represents the commutator of A and B .

This inequality form of the Uncertainty Principle is not an arbitrary construct; it stems from the mathematical foundation provided by the Cauchy-Schwarz inequality: $|\langle u|v \rangle|^2 \leq \langle u|u \rangle \langle v|v \rangle$.

Although the underlying mathematics are intriguing, we will not delve into those details in this text.

1.7 *Further Readings on Quantum Mechanics

For students and researchers aiming to deepen their understanding of quantum mechanics, we recommend the following textbooks.

- Junichiro Kono. *Quantum Mechanics for Tomorrow's Engineers: New Edition*. English. New edition. Cambridge University Press, Sept. 29, 2022, page 350. ISBN: 978-1108842587.

This text presents a pioneering approach to quantum mechanics, tailored for students in quantum computing and engineering. This comprehensive undergraduate textbook bridges the gap between abstract quantum principles and their tangible applications in the field of quantum engineering.

Kono's work is unique in its focus on current technologies, steering clear of historical methodologies to prepare students for contemporary engineering challenges. The book offers a robust introduction to quantum information fundamentals and demonstrates their practicality through a plethora of real-world examples. These include quantum well infrared photodetectors, solar cells, and quantum teleportation, as well as cutting-edge topics like quantum computing, and band gap engineering.

- Claude Cohen-Tannoudji, Bernard Diu, and Franck Laloë. *Quantum Mechanics, Vol 1: Basic Concepts, Tools, and Applications*. Wiley, 2019. ISBN: 978-3-527-34553-3.

The series of three volumes, authored by Nobel Prize laureate Claude Cohen-Tannoudji and colleagues, stands as a seminal contribution to the field of quantum mechanics. This comprehensive series not only introduces the fundamentals but also delves into more intricate aspects of the subject.

The first volume is particularly noteworthy for its excellent introductory content. It provides a clear and detailed exposition of the fundamental postulates of quantum mechanics. The text excels in going beyond simple definitions, integrating rich discussions and intuitive explanations that shed light on complex concepts. A remarkable aspect of this volume is its focus on two-level systems. These systems form the cornerstone of qubit systems in quantum computing and quantum information science, and the dedicated sections

offer valuable insights directly relevant to current research and technological advancements in these areas.

The series continues to offer in-depth knowledge in its subsequent volumes. Volume 2, titled “Angular Momentum, Spin, and Approximation Methods,” and Volume 3, “Fermions, Bosons, Photons, Correlations, and Entanglement,” extend the discussion to more advanced topics. These volumes provide a comprehensive understanding of the subject, making the series an indispensable resource for anyone keen to explore the depths of quantum mechanics.

- Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *Quantum Mechanics (Feynman Lectures on Physics, Volume 3)*. Basic Books, Oct. 4, 2011. ISBN: 978-0465025015.

This volume, part of the renowned “Feynman Lectures on Physics,” provides an inspirational and insightful exploration of quantum mechanics. Authored by the legendary physicist Richard P. Feynman, this book is celebrated for its ability to convey complex ideas in an accessible manner. Feynman’s unique teaching style combines rigorous scientific understanding with a deep appreciation of the beauty underlying quantum phenomena. This text is an excellent resource for anyone seeking a deeper conceptual understanding of quantum mechanics, presented in a way that is both engaging and thought-provoking.

- J. J. Sakurai and Jim Napolitano. *Modern Quantum Mechanics*. Cambridge University Press, 2017. ISBN: 978-1-108-47322-4.

This text offers a comprehensive and detailed examination of advanced quantum mechanics concepts and techniques, particularly useful for those involved in quantum simulation. The book adeptly covers a range of advanced topics including Symmetry in Quantum Mechanics, Scattering Theory, Approximation Methods, Identical Particles, and Relativistic Quantum Mechanics. Its clear and structured approach makes complex ideas more accessible, while providing practical insights into how these principles are applied in modern quantum research and technology.

1.8 Topic Reviews

In this chapter, we navigated through the intricate landscape of quantum mechanics using photons as our guide. We delved into a myriad of core concepts and mathematical formulations, serving both as an introduction and a focused exploration into the quantum realm. To aid in assimilating this rich tapestry of knowledge, this section provides a concise review of the essential concepts and equations introduced.

1.8.1 Review of Postulates 1-3

- **Postulate 1: Quantum State**
 - States of quantum systems are described by normalized complex vectors in a Hilbert space.
- **Implications from Postulate 1**
 - Orthogonal states are mutually exclusive, rendering them orthonormal.
 - A general quantum state can exist as a superposition of basis states.

- The global phase of a state vector has no physical significance.

- **Postulate 2: Quantum Observables**

- Physically measurable quantities correspond to Hermitian operators.
- Measurement outcomes and states are defined by the eigenvalues and eigenvectors of these operators.

- **Implications from Postulate 2**

- Quantum observables often have discrete eigenvalues, leading to quantization of measurement outcomes.
- The choice of observable specifies the basis in which the measurement will be expressed.
- Dirac bra-ket notation universally represents quantum states and operators in a basis-independent manner.

- **Postulate 3: Quantum Measurements**

- Measurement outcomes and probabilities are governed by the Born rule.
- Measurements induce a collapse of the quantum state.

- **Implications from Postulate 3**

- Quantum measurements are inherently probabilistic.
- A measurement irreversibly alters the state.
- Subsequent measurements yield identical results.
- Statistical average for measuring an observable can be calculated.
- The Born rule connects quantum states to measurement probabilities.
- Probabilities of all possible outcomes must sum to 1.

- **Cross-Postulate Implications**

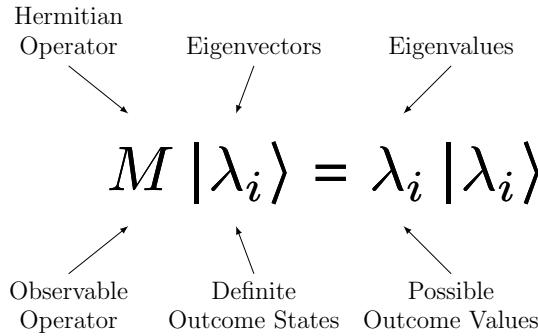
- Quantum Interference: A phenomenon that arises due to the superposition of states and subsequent measurements.
- The Uncertainty Principle: Limitations exist for the simultaneous measurement of non-commuting observables.

See Table 1.2: Summary of Photon Polarization States and Observables

1.8.2 Review of Photons as Quantum Systems

- **Photon as Quantum Systems**

- Photons are elementary particles and the quanta of electromagnetic fields, including light.
- Photons are vital in various quantum technologies such as lasers, quantum cryptography, and quantum computing.
- The polarization state of a photon can be represented as a normalized complex vector in a Hilbert space.



General superposition state: $|\psi\rangle = \sum_i c_i |\lambda_i\rangle$ with $\langle\psi|\psi\rangle = \sum_i |c_i|^2 = 1$.

Probability of outcome λ_i : $|\langle\lambda_i|\psi\rangle|^2 = |c_i|^2$. Post-measurement state: $|\lambda_i\rangle$.

Expected value (statistical average) of M : $\langle M \rangle = \sum_i |c_i|^2 \lambda_i = \langle\psi|M|\psi\rangle$.

Figure 1.13: Summary of the Observable and Measurement Postulates

- Three essential sets of orthogonal photon polarization states: rectilinear, diagonal, and circular.

- **Characteristics of Photons**

- Polarization states that are orthogonal are mutually exclusive, thus aligning with the orthonormal property in quantum mechanics.
- A photon's polarization can exist as a superposition of basis polarization states.
- Quantum polarization states can also be superposed and entangled, giving rise to distinctly quantum behaviors with no classical analogs.
- Quantum interference is observed when photons are sent through two-path systems, supporting the superposition principle.

- **Photon Observables**

- Polarization of a photon serves as an example of a quantum observable.
- Measurements related to polarization can be defined through appropriate Hermitian operators.

- **Photon Measurements**

- Measurement of a photon's polarization yields results consistent with the Born rule.
- After measurement, the photon state collapses to the eigenstate corresponding to the measured polarization angle.

- **Characteristics of Photon Measurements**

- Similar to other quantum systems, photon measurements are probabilistic.

- The act of measuring the polarization collapses the photon's state, altering it irreversibly.
- Repeated measurements on the same photon would yield the same polarization angle, provided the photon is not altered between measurements.

See **Table 1.2:** Summary of Photon Polarization States and Observables

1.9 Summary and Conclusions

Foundational Principles

In this chapter, we have delved into the foundational principles that govern quantum computing and quantum information science (QCI). Our discussion emphasized the importance of understanding these principles, not just for theoretical clarity but also for practical applications.

We began by examining quantum states, described by normalized complex vectors in a vector space, setting the stage for the introduction of quantum bits or qubits. We explored the critical concept of superposition, which underlies the parallel computational capabilities unique to quantum systems.

We then transitioned to the discussion of quantum observables, represented by Hermitian operators. These operators define physically measurable quantities, and their eigenvalues and eigenvectors are intimately related to the outcomes and probabilities of quantum measurements.

Next, we addressed the essential rules governing quantum measurements, focusing on the Born rule for calculating probabilities and the concept of state collapse post-measurement. We underscored the inherent probabilistic nature of quantum mechanics and the irreversible effect of measurements on quantum states.

Further, we discussed ancillary but important concepts such as quantum interference, the uncertainty principle, and the lack of physical significance for the global phase of a quantum state. These discussions enriched our understanding of the subtleties and complexities involved in quantum mechanics and quantum computation.

General Framework

While our approach has been tailored towards two-level quantum systems like qubits, the principles discussed are general and applicable to higher-dimensional quantum systems as well. Throughout the chapter, we have presented examples using photons to make the concepts more relatable, without compromising on academic rigor.

Our framework positions QCI as an interdisciplinary field, combining insights from quantum physics, computer science, and mathematics. This multidisciplinary lens offers a more comprehensive understanding, preparing the reader for the rich and diverse landscape of quantum computing and information.

Upcoming Topics

In the upcoming chapters, we will initially focus on half-integer spins, another physical realization of a qubit system. Following this, we will introduce a general theoretical framework for qubits. Subsequently, we will explore the remaining

principles of Quantum Evolution and Quantum Composition, further deepening our understanding of the theory and practice of Quantum Computing and Information.

Problem Set 1

1.1 Compare Eqs. 1.12 and 1.16, and derive formulas for

- (a) γ and δ in terms of μ and ν
- (b) μ and ν in terms of γ and δ

using both matrix representation and Dirac notation.

1.2 Show that when $\phi = \frac{\pi}{2}$, Eq. 1.19 represents an elliptical polarization state with major radius aligned with $|V\rangle$ or $|H\rangle$. What are its major and minor radii?

1.3 Investigate the relationship between the tilt angle φ , and the major and minor radii of the ellipse in Fig. 1.5, with respect to θ and ϕ .

1.4 Elaborate the conditions for a , c , and d , such that

$$\begin{bmatrix} a & \sin \theta e^{i\phi} + i \cos \theta e^{-i\phi} \\ c & d \end{bmatrix}$$

is a Hermitian matrix, where $\phi, \theta \in \mathbb{R}$ and $a, c, d \in \mathbb{C}$.

1.5 A general linear polarization state $|\theta\rangle$ at angle θ relative to $|V\rangle$, and its orthogonal state $|\theta_\perp\rangle$, are given by

$$|\theta\rangle = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \quad |\theta_\perp\rangle = \begin{bmatrix} \sin \theta \\ -\cos \theta \end{bmatrix}.$$

Derive the corresponding observable operator M_θ .

1.6 Derive the observable operator M_P for the general (elliptical) polarization state given in Eq. 1.19.

1.7 Consider photons polarized at an angle α relative to $|V\rangle$, and measure them using a polarizer oriented at an angle β . Show that the statistical average of the measurement is $\cos^2(\alpha - \beta)$, which depends only on the relative angle $\alpha - \beta$.

1.8 Elaborate the conditions of α, β, δ and γ under which measuring photons against the general polarization state $\alpha|V\rangle + \beta|H\rangle$ is compatible with measuring against another general polarization state $\delta|V\rangle + \gamma|H\rangle$.

2. Fundamentals of Spin Systems

Contents

2.1	Spin, Angular Momentum, and Magnetic Moment	42
2.1.1	Fermions, Bosons, and Anyons	42
2.1.2	1/2 Spin	42
2.1.3	Magnetic Moment	43
2.2	Spin-1/2 States and Pauli Matrices	43
2.2.1	Spin-1/2 States and Observables	43
2.2.2	Properties of Pauli Matrices	44
2.3	* General Spin State Representation	45
2.3.1	* Defining the Spin Observable for Any Direction	45
2.3.2	* Spin State in Any Direction	46
2.3.3	General Spin State	47
2.3.4	Spin State and Polarization State Correspondence	47
2.4	The Bloch Sphere	48
2.4.1	Key States on the Bloch Sphere	49
2.4.2	Representation of Pauli X, Y, and Z on the Bloch Sphere	49
2.5	Spin Measurement	49
2.5.1	The Stern-Gerlach Experiment	50
2.5.2	Cascaded Stern-Gerlach Experiments	52
2.5.3	* Modern Spin Experiments	52
2.6	Summary and Conclusions	53
	Problem Set 2	54

In this chapter, we explore the fundamental principles of quantum mechanics as they pertain to spin systems. Understanding the intricacies of spin is vital, particularly for fields that heavily rely on quantum mechanics, such as quantum computing. Along the way, we will introduce two essential tools: the Pauli matrices and the Bloch sphere.

2.1 Spin, Angular Momentum, and Magnetic Moment

In classical physics, objects like a spinning tennis ball possess angular momentum. This angular momentum is a vector quantity directed along the axis of rotation, and its magnitude can take continuous values. However, for quantum particles such as electrons, protons, and neutrons, as well as composite particles like atoms, the scenario is fundamentally different. The angular momentum of these particles is quantized and can only take discrete values. Moreover, these particles can have a permanent, non-zero angular momentum, as if they are perpetually spinning. This intrinsic form of angular momentum is known as *spin*.

Spin is a pervasive phenomenon in physics with a plethora of applications, including but not limited to electron microscopy, magnetic resonance, and quantum computation.

2.1.1 Fermions, Bosons, and Anyons

Particles can be classified into two broad categories based on their spin: fermions and bosons. Fermions have half-integer spins ($\frac{1}{2}, \frac{3}{2}, \dots$) while bosons possess integer spins (0, 1, 2, ...). Essentially all particles of conventional matter in our observable universe, including electrons, protons, and neutrons, are fermions. Photons, which are particles of light, have an integer spin of 1, classifying them as bosons. This distinction between integer and half-integer spins arises from the principles of special relativity, which dictate how particles' wave functions transform under the Lorentz group.

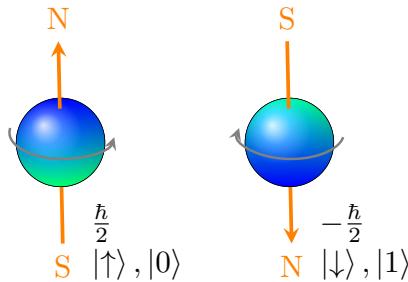
In addition to fermions and bosons, there is another intriguing class of quasiparticles known as anyons. Unlike fermions and bosons, which adhere strictly to integer or half-integer spin values, anyons exhibit more exotic behaviors in two-dimensional systems. Their spins can take on a range of values between integers and half-integers, leading to unconventional quantum statistics. Anyons are currently of great interest, especially in the burgeoning field of topological quantum computing, where they are considered as potential carriers of quantum information.

2.1.2 1/2 Spin

Particles with a spin of $\frac{1}{2}$, such as electrons and some atomic nuclei, are of particular interest in the domain of quantum computing. As illustrated in Fig. 2.1, these particles exhibit two discrete quantum states, commonly termed as up and down states. These are mathematically represented by $|\uparrow\rangle$ and $|\downarrow\rangle$, or equivalently $|0\rangle$ and $|1\rangle$. The angular momentum corresponding to these states is $S_u = \frac{\hbar}{2}$ and $S_d = -\frac{\hbar}{2}$, where \hbar is the reduced Planck constant.

In addition to serving as two-level quantum systems that can act as qubits, the mathematical framework that describes spin- $\frac{1}{2}$ systems can easily be generalized to model qubits, forming a cornerstone for quantum computing algorithms.

One fascinating aspect of spin- $\frac{1}{2}$ particles is their unusual rotational behavior. Unlike classical objects that return to their original state after a 360° rotation, spin- $\frac{1}{2}$ particles require a 720° rotation for their wavefunction to return to its original state. This intriguing property, fundamentally linked to the topology of quantum states, significantly departs from classical expectations.

**Figure 2.1:** 1/2 Spin

Given its significance in quantum computing, our subsequent discussions will concentrate primarily on spin- $\frac{1}{2}$ systems.

2.1.3 Magnetic Moment

For charged quantum particles like electrons and protons, the intrinsic spin is associated with a magnetic moment. Mathematically, this magnetic moment is proportional to the spin, denoted as $\vec{\mu} \propto \vec{S}$. This relationship is fundamental to technologies that exploit magnetic interactions, such as Magnetic Resonance Imaging (MRI), quantum computing, and quantum sensing.

2.2 Spin-1/2 States and Pauli Matrices

2.2.1 Spin-1/2 States and Observables

Spin, as an intrinsic form of angular momentum, is represented by a vector in three-dimensional space. For particles with spin- $\frac{1}{2}$, this vector assumes two quantized values, $\frac{\hbar}{2}$ and $-\frac{\hbar}{2}$, in any chosen direction, where \hbar is the reduced Planck constant. In the z -direction, these states are typically referred to as up and down, denoted as $|0\rangle \equiv |\uparrow\rangle$ and $|1\rangle \equiv |\downarrow\rangle$, which constitute the computational basis. Mathematically, these states are defined as:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (2.1)$$

The observable S_z , associated with measuring spin in the z -direction, has these states as its eigenstates, and $\frac{\hbar}{2}$ and $-\frac{\hbar}{2}$ as the corresponding eigenvalues. S_z can be represented through its spectral expansion as:

$$S_z = \frac{\hbar}{2} |0\rangle\langle 0| - \frac{\hbar}{2} |1\rangle\langle 1| = \frac{\hbar}{2} Z, \quad (2.2)$$

where Z is the Pauli Z matrix, represented as

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (2.3)$$

In quantum computing contexts, we often work directly with Z instead of S_z , given that they differ only by a factor of $\frac{\hbar}{2}$.

Similarly, in the x -axis, the state vectors are expressed as:

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (2.4a)$$

$$|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \quad (2.4b)$$

The associated observable is the Pauli X matrix:

$$X = |+\rangle\langle+| - |-\rangle\langle-| = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (2.5)$$

In the y -axis, the state vectors are:

$$|+_i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix}, \quad (2.6a)$$

$$|-_i\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix}. \quad (2.6b)$$

$|+_i\rangle$ and $|-_i\rangle$ are often written as $|i_+\rangle$ and $|i_-\rangle$, or $|i\rangle$ and $|-i\rangle$ in other texts.

The associated observable is the Pauli Y matrix:

$$Y = |+_i\rangle\langle+_i| - |-_i\rangle\langle-_i| = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}. \quad (2.7)$$

Exercise 2.1 Derive the matrix forms of Pauli matrices X and Y based on their bra-ket representations.

The specific forms of these state vectors are not arbitrary but are the result of insightful contributions from physicists, notably Pauli. These forms will be further justified in the subsequent discussion about general spin directions.

2.2.2 Properties of Pauli Matrices

Pauli matrices serve functions beyond acting as spin observables; they play a pivotal role in various domains of quantum computing and quantum physics. As such, a thorough understanding of their properties is crucial. These properties are highlighted below, with more details in Appendix D. (For succinct presentation of certain relations, Pauli matrices are often denoted as $\sigma_1 \equiv X$, $\sigma_2 \equiv Y$, $\sigma_3 \equiv Z$. In the following, we use σ to represent any one of them.)

1. Hermitian Property: $\sigma^\dagger = \sigma$.
2. Unitary: $\sigma^\dagger \sigma = I$.
3. Involutory Property: $\sigma^2 = I$.
4. Commutation Relation: $[\sigma_j, \sigma_k] = 2i\varepsilon_{jkl}\sigma_l$, where ε_{jkl} is the Levi-Civita permutation symbol.

5. Anti-Commutation Relations: $\{\sigma_i, \sigma_j\} = 2\delta_{ij}I$, where δ_{ij} is the Kronecker delta function.
6. Algebraic Completeness: Pauli matrices (together with I) provides a complete basis for general matrices in $\mathbb{C}^{2 \times 2}$.
7. Spin Angular Momentum: Pauli matrices represent the spin angular momentum operator in the x , y , and z directions with the correct rotation properties.

Why Complex Numbers?

One may wonder why one of the Pauli matrices contains i . Is it fundamentally necessary or just a matter of convenience? It turns out the presence of i is necessary for the Pauli operators:

1. to satisfy anti-commutation relation.
2. to generate the correct rotation properties of angular momentum for spin- $\frac{1}{2}$ particles.
3. to provide a complete, orthogonal basis for 2×2 Hermitian matrices.

Exercise 2.2 Given the eigenvectors $|0\rangle$ and $|1\rangle$ of Pauli matrix Z , $|+\rangle$ and $|-\rangle$ of X , and $|+_i\rangle$ and $|-_i\rangle$ of Y , prove that

$$\begin{aligned}\langle 0|Z|0\rangle &= 1, & \langle 1|Z|1\rangle &= -1, \\ \langle +|X|+\rangle &= 1, & \langle -|X|-\rangle &= -1, \\ \langle +_i|Y|+_i\rangle &= 1, & \langle -_i|Y|-_i\rangle &= -1,\end{aligned}$$

and

$$\begin{aligned}\langle 0|X|0\rangle &= \langle 1|X|1\rangle &= \langle 0|Y|0\rangle &= \langle 1|Y|1\rangle &= 0, \\ \langle +|Y|+\rangle &= \langle -|Y|-\rangle &= \langle +|Z|+\rangle &= \langle -|Z|-\rangle &= 0, \\ \langle +_i|Z|+_i\rangle &= \langle -_i|Z|-_i\rangle &= \langle +_i|X|+_i\rangle &= \langle -_i|X|-_i\rangle &= 0.\end{aligned}$$

Provide a physical interpretation of these results in terms of measuring the X , Y , and Z components of spins oriented in various directions.

2.3 *General Spin State Representation

In this subsection, we explore the representation of spin states and observables along a general direction in three-dimensional space. This direction is specified by \hat{u} , a real unit vector with Cartesian components u_x , u_y , and u_z , where $u_x^2 + u_y^2 + u_z^2 = 1$. In spherical coordinates (see Fig. 2.2), \hat{u} is described by the polar angle θ and azimuthal angle ϕ :

$$\hat{u} = \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{bmatrix}. \quad (2.8)$$

2.3.1 *Defining the Spin Observable for Any Direction

The Pauli matrices X , Y , and Z represent the Cartesian components of the observable of an angular momentum, a three-dimensional vector. Consequently, the spin

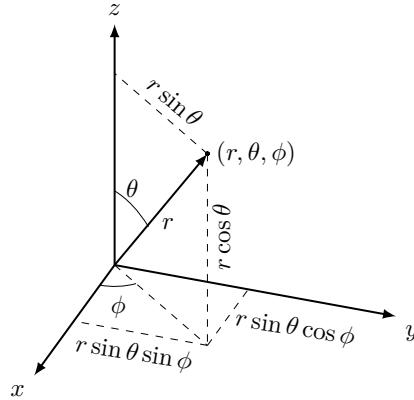


Figure 2.2: Spherical Coordinate System

observable along the direction of \hat{u} can be expressed as:

$$S = u_x X + u_y Y + u_z Z = \sin \theta \cos \phi X + \sin \theta \sin \phi Y + \cos \theta Z. \quad (2.9)$$

Substituting X , Y , and Z with their respective matrices, we find:

$$S \equiv \sigma_u = \begin{bmatrix} \cos \theta & e^{-i\phi} \sin \theta \\ e^{i\phi} \sin \theta & -\cos \theta \end{bmatrix}. \quad (2.10)$$

To validate the definition of S in Eq. 2.9, the expected value $\langle S \rangle$ (see § 1.5.3) for a spin state should align with the Cartesian components of \hat{u} :

$$\langle +|S|+ \rangle = u_x = \sin \theta \cos \phi, \quad \langle -|S|- \rangle = -u_x, \quad (2.11a)$$

$$\langle +_i|S|+_i \rangle = u_y = \sin \theta \sin \phi, \quad \langle -_i|S|-_i \rangle = -u_y, \quad (2.11b)$$

$$\langle 0|S|0 \rangle = u_z = \cos \theta, \quad \langle 1|S|1 \rangle = -u_z. \quad (2.11c)$$

These relations can be verified through the bra-ket notation using the results from Exercise 2.2 or by directly computing them in matrix form.

Exercise 2.3 Verify Eq. 2.11.

2.3.2 * Spin State in Any Direction

The eigenvectors of S in Eq. 2.10 are the spin states in the \hat{u} and $-\hat{u}$ directions:

$$|+_u\rangle = \cos \frac{\theta}{2} |0\rangle + \sin \frac{\theta}{2} e^{i\phi} |1\rangle = \begin{bmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\phi} \end{bmatrix}, \quad (2.12a)$$

$$|-_u\rangle = -\sin \frac{\theta}{2} |0\rangle + \cos \frac{\theta}{2} e^{i\phi} |1\rangle = \begin{bmatrix} -\sin \frac{\theta}{2} \\ \cos \frac{\theta}{2} e^{i\phi} \end{bmatrix}. \quad (2.12b)$$

Exercise 2.4 Verify that $S|+u\rangle = |+u\rangle$ and $S|-u\rangle = -|-u\rangle$.

Exercise 2.5 You may find $|+u\rangle$ and $|-u\rangle$ expressed differently in other texts:

$$|+u\rangle = \cos \frac{\theta}{2} e^{-i\frac{\phi}{2}} |0\rangle + \sin \frac{\theta}{2} e^{i\frac{\phi}{2}} |1\rangle = \begin{bmatrix} \cos \frac{\theta}{2} e^{-i\frac{\phi}{2}} \\ \sin \frac{\theta}{2} e^{i\frac{\phi}{2}} \end{bmatrix}, \quad (2.13a)$$

$$|-u\rangle = -\sin \frac{\theta}{2} e^{-i\frac{\phi}{2}} |0\rangle + \cos \frac{\theta}{2} e^{i\frac{\phi}{2}} |1\rangle = \begin{bmatrix} -\sin \frac{\theta}{2} e^{-i\frac{\phi}{2}} \\ \cos \frac{\theta}{2} e^{i\frac{\phi}{2}} \end{bmatrix}. \quad (2.13b)$$

Explain why this definition of $|+u\rangle$ and $|-u\rangle$ is equivalent to the one in Eq. 2.12.

2.3.3 General Spin State

The state $|+u\rangle$ serves as a general representation of a spin state. In fact, all the specialized states we discussed earlier, such as $|0\rangle$ and $|+\rangle$, including $|-u\rangle$, can be viewed as special cases of $|+u\rangle$. This is summarized in Table 2.1.

Spin State	θ	ϕ
$ 0\rangle$	0	0
$ 1\rangle$	π	0
$ +\rangle$	$\pi/2$	0
$ -\rangle$	$-\pi/2$	0
$ +i\rangle$	$\pi/2$	$\pi/2$
$ -_i\rangle$	$\pi/2$	$-\pi/2$
$ -u\rangle$	$\theta + \pi$	ϕ

Table 2.1: Special Spin States

Exercise 2.6 Validate that the expression for $|+u\rangle$ in Eq. 2.12 reproduces the special cases listed in Table 2.1.

2.3.4 Spin State and Polarization State Correspondence

Even though spin and photon polarization are two distinct physical phenomena, they share many mathematical similarities because they both describe two-level quantum systems. In fact, various spin states can be mapped to photon polarization states, as listed in Table 2.2. The corresponding observables manifest as different matrices, because we assigned 1, 0 for photon polarization, while 1, -1 for spin states.

In this section, we extended the representation of spin states and spin observables to an arbitrary direction in three-dimensional space. Understanding these general representations sets the foundation for studying more advanced topics in quantum mechanics.

Spin State	Spin Observable	Polarization State	Polarization Observable
$ 0\rangle$	Z	$ V\rangle$	M_V
$ 1\rangle$	Z	$ H\rangle$	M_H
$ +\rangle$	X	$ D\rangle$	M_D
$ -\rangle$	X	$ A\rangle$	M_A
$ +_i\rangle$	Y	$ L\rangle$	M_L
$ -_i\rangle$	Y	$ R\rangle$	M_R
$ +_u\rangle$	S	$ P\rangle$	M_P

(See Table 1.2 for definitions of polarization states and observables, and Eq. 1.19 for $|P\rangle$)

Table 2.2: Spin State and Polarization State Correspondence

2.4 The Bloch Sphere

Recalling the expression for the generic spin state $|+_u\rangle$ from Eq. 2.12, we find that each state is specified by the polar angle θ and azimuthal angle ϕ . It is then natural to map each point on the surface of a unit sphere to a corresponding spin state $|+_u\rangle$. Conversely, each possible spin state can also be represented as a point on this sphere. This sphere, known as the Bloch Sphere, serves as a geometrical representation for any qubit state $|\psi\rangle$, as shown in Fig. 2.3.

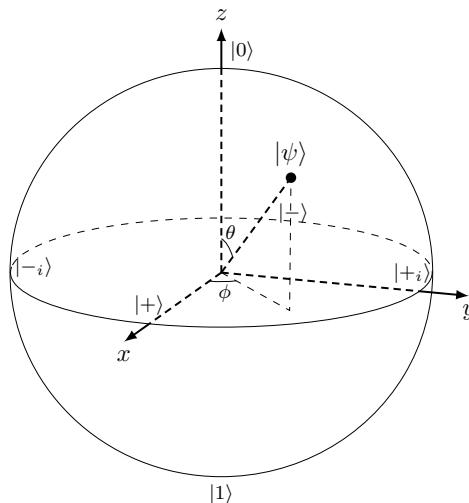


Figure 2.3: Key Points on the Bloch Sphere

Bloch Sphere

The Bloch sphere is an intuitive geometrical representation that maps any qubit

state $|\psi\rangle$ to a point on a unit sphere. The state $|\psi\rangle$ is parameterized by the polar angle θ and azimuthal angle ϕ :

$$|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + \sin \frac{\theta}{2} e^{i\phi} |1\rangle = \begin{bmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\phi} \end{bmatrix}. \quad (2.14)$$

The expression for $|\psi\rangle$ in Eq. 2.14 is the same as for $|+_u\rangle$ but applies universally to any qubit state.

Note that we typically restrict $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$, recognizing that states $(-\theta, \phi + \pi)$ and $(-\theta, \phi - \pi)$ are equivalent to (θ, ϕ) .

2.4.1 Key States on the Bloch Sphere

As depicted in Fig. 2.3, certain key states correspond to specific points on the Bloch sphere. Specifically, the north and south poles along the z -axis are associated with the $|0\rangle$ and $|1\rangle$ states, respectively. Eigenstates of the Pauli X operator, namely $|+\rangle$ and $|-\rangle$, map to points along the x -axis, while eigenstates of the Pauli Y operator, $|+_i\rangle$ and $|-_i\rangle$, align along the y -axis.

It is worth noting that orthogonal states are positioned at antipodal points on the sphere, separated by an angle of 180° , rather than the 90° one might intuitively expect. For instance, the orthogonal complement of a state $|\psi\rangle$ represented by (θ, ϕ) is found at $(\pi - \theta, \pi + \phi)$ or $(\theta + \pi, \phi)$, up to a global phase of -1 . This arrangement reflects the intriguing property of quantum spin, where a full rotation requires a 720° turn, not the classical 360° .

2.4.2 Representation of Pauli X, Y, and Z on the Bloch Sphere

The Pauli operators X , Y , and Z induce specific rotations on the Bloch sphere. For instance, the action of X on a state $|\psi\rangle$, transforming $|\psi\rangle$ to $X|\psi\rangle$, corresponds to a 180° rotation about the x -axis. Similarly, applying Y and Z corresponds to 180° rotations about the y -axis and z -axis, respectively. These rotations are illustrated in Fig. 2.4.

Note: A 180° rotation on the Bloch sphere is associated with the action of the Pauli operators X , Y , or Z . Technically, this rotation is equivalent to applying a Pauli operator up to a global phase factor. However, such a global phase factor does not change the observable properties of a quantum state.

On the Bloch sphere, the effects of these rotations manifest as flips between certain basis states. Specifically, X swaps the states $|0\rangle$ and $|1\rangle$, and $|+_i\rangle$ and $|-_i\rangle$; Z swaps $|+\rangle$ and $|-\rangle$, and $|+_i\rangle$ and $|-_i\rangle$; while Y swaps $|0\rangle$ and $|1\rangle$, and $|+\rangle$ and $|-\rangle$.

Exercise 2.7 Verify the above transformations using bra-ket algebra, for example, confirm that $X|+_i\rangle = |-_i\rangle$, $Y|+\rangle = |-\rangle$, and $Z|-\rangle = |+\rangle$.

2.5 Spin Measurement

In this section, our focus shifts to the measurement of spins. While modern methods

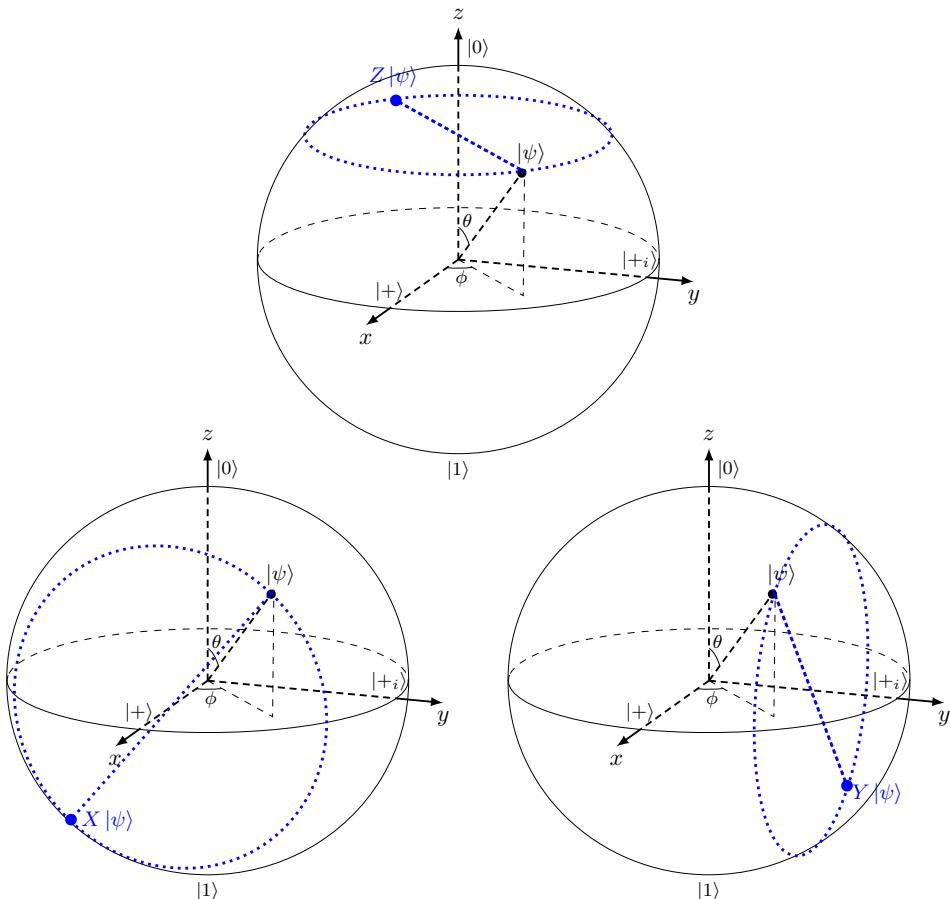


Figure 2.4: Pauli X , Y , and Z as Rotations on the Bloch Sphere

are designed to suit various quantum computing platforms, the Stern-Gerlach (SG) experiment serves as a canonical approach for understanding spin measurement. This experiment provides vital insights into the quantization of angular momentum and serves as a foundational framework for more advanced measurement techniques. We begin by examining the key aspects of the SG experiment.

2.5.1 The Stern-Gerlach Experiment

The Stern-Gerlach experiment, a cornerstone in the study of particle spin, is predicated on directing a beam of particles, such as electrons or atoms, through an inhomogeneous magnetic field (Fig. 2.5). Given the magnetic moments of these particles (§ 2.1.3), their different spin orientations cause them to deviate along the magnetic field gradient and register on a detection screen.

As depicted, the experiment primarily measures the Z observable, due to the magnetic field gradient being mainly along the z -axis. The observable Z has two eigenvalues, which correspond to the spin-up and spin-down states. While the initial spin state of the particles is random, the measurement collapses it to either spin-up or spin-down. Therefore, two distinct spots appear on the screen, in stark contrast to the continuous distribution one would expect for classical angular momentum.

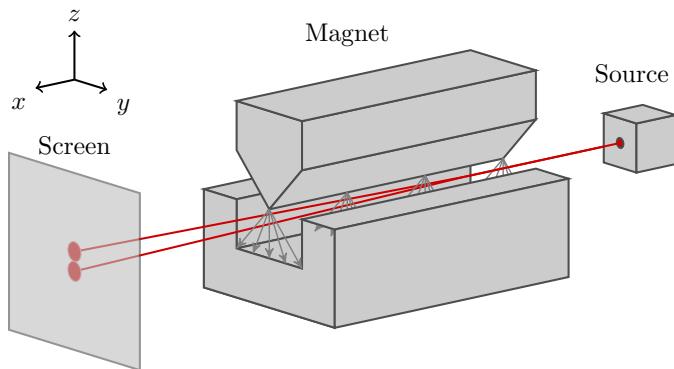


Figure 2.5: Stern-Gerlach Experiment

The original experiment, carried out by Otto Stern and Walther Gerlach in 1922 using silver atoms, conclusively revealed two separate spots, thus providing unequivocal evidence for the quantization of angular momentum and magnetic moment. This groundbreaking discovery laid the essential groundwork for the development of quantum theories.



The necessity of a magnetic field gradient in the SG experiment can be attributed to the dipole nature of the magnetic moment, which has the same orientation as the spin. In a uniform field, the potential energy depends on the orientation but is position-independent. A gradient in the magnetic field thus gives rise to a force that causes the observed deflection.

A Fascinating Story

Stern described an early episode (Source: Physics Today):

“After venting to release the vacuum, Gerlach removed the detector flange. But he could see no trace of the silver atom beam and handed the flange to me. With Gerlach looking over my shoulder as I peered closely at the plate, we were surprised to see gradually emerge the trace of the beam. ... Finally, we realized what [had happened]. I was then the equivalent of an assistant professor. My salary was too low to afford good cigars, so I smoked bad cigars. These had a lot of sulfur in them, so my breath on the plate turned the silver into silver sulfide, which is jet black, so easily visible. It was like developing a photographic film.”

The anecdote underscores a significant point about scientific research: sometimes, the path to discovery is not straightforward and may even be filled with seemingly trivial or random events that turn out to be pivotal. It serves as a testament to the resilience and ingenuity that are often required in the scientific journey. It is not just high-tech equipment or sophisticated theories that make for great science, but also an open mind, attention to detail, and the ability to adapt and learn from unexpected situations. In some instances, as comically illustrated by Stern’s account, the constraints or limitations we face might inadvertently lead us to meaningful insights, reaffirming the adage that “necessity is the mother of invention.”

2.5.2 Cascaded Stern-Gerlach Experiments

The Measurement Postulate can be well demonstrated using cascaded Stern-Gerlach experiments, as illustrated in Fig. 2.6. These cascaded setups are conceptually analogous to the cross-polarizer experiments detailed in § 1.5.2.

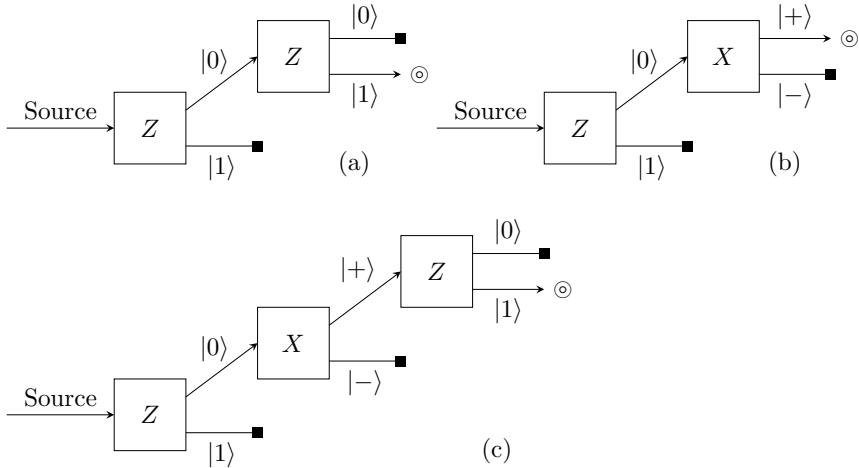


Figure 2.6: Cascaded Stern-Gerlach Experiments

In part (a), a beam of particles with random spin orientations is directed through a Stern-Gerlach apparatus (SGA) oriented along the z -axis. This first SGA measures the spin with respect to the Z observable and bifurcates the beam into two separate paths, each corresponding to one of Z 's eigenstates, $|0\rangle$ and $|1\rangle$, in a $50 : 50$ ratio. The $|0\rangle$ beam is then directed to a second SGA, also oriented along the z -axis. At this stage, no particles will be detected in the $|1\rangle$ output, confirming that the first measurement collapsed the spin state to $|0\rangle$, which is orthogonal to $|1\rangle$.

Part (b) varies from part (a) by having the second SGA oriented along the x -axis. The measurement outcomes are $|+\rangle$ and $|-\rangle$, the eigenstates of the X observable. Given that $|\langle 0|+ \rangle|^2 = |\langle 0|- \rangle|^2 = 0.5$, approximately 25% of the particles will be found in the $|+\rangle$ state.

In part (c), an additional SGA oriented in the x -axis is inserted between the two Z -oriented SGAs. Now, about 12.5% of particles are observed in the $|1\rangle$ output of the final SGA. This is because the middle X -oriented SGA forces the spins into $|+\rangle$ and $|-\rangle$ states, which then feed into the final Z -oriented SGA. This SGA subsequently splits the $|+\rangle$ state into $|0\rangle$ and $|1\rangle$ with a $50 : 50$ ratio. The net probability of finding a particle in the $|1\rangle$ state is $0.5 \times 0.5 \times 0.5 = 0.125$ or 12.5%, mirroring the behavior seen in the cross-polarizer experiments.

2.5.3 *Modern Spin Experiments

In quantum mechanics, measurement refers to the extraction of classical information from a quantum system. While early experiments such as Stern-Gerlach provided initial insights into quantum spin, contemporary approaches have significantly evolved, particularly with the advent of quantum computing. One of the leading

candidates for building qubits in a quantum computer is the use of spin-based qubits, which often rely on more advanced techniques for precise spin manipulation and measurement [58].

Spin-Based Qubits

Spin-based qubits are typically built using semiconductor quantum dots or trapped ions. The benefit of using these systems is that they offer a higher level of control and scalability. The challenge lies in accurately manipulating and measuring the spin states without undesirable interactions, which would disrupt the quantum information stored in the qubits.

Spin Precession and Rabi Oscillations

Two fundamental phenomena often employed in modern spin experiments are spin precession and Rabi oscillations. Spin precession refers to the motion of a quantum spin vector around an external magnetic field, much like the way a gyroscope precesses around an axis. Rabi oscillations involve driving transitions between different quantum states using an oscillating external field, commonly a microwave field. We will study these two phenomena in §§ 4.5.1 and 4.5.2.

Both phenomena are invaluable for spin manipulation in quantum computing. Spin precession allows for the precise control of qubit states by modulating the external magnetic field. On the other hand, Rabi oscillations are essential for implementing quantum gates, which are the basic operations in quantum computing algorithms.

Measurement Techniques

When a spin-1/2 particle, such as an electron, is measured along a specific axis, the orientation of its spin relative to that axis determines the outcome of the measurement. If the spin state is aligned with the measurement axis, the result is deterministic, yielding a definitive spin value (up or down along that axis). This principle underlies many modern spin experiments and is essential for confirming the theoretical predictions of quantum mechanics.

Advanced measurement techniques such as spin resonance and quantum non-demolition measurements are also integral to the robustness of modern spin experiments. These methods allow for the highly sensitive detection of spin states while minimizing the disturbance to the quantum system, thus enabling the extraction of reliable classical information.

Modern spin experiments have come a long way from their early counterparts, employing sophisticated methods for spin manipulation and measurement. These developments not only deepen our understanding of quantum mechanics but also pave the way for practical applications like quantum computing.

2.6 Summary and Conclusions

In this chapter, we laid the groundwork by introducing the essential elements of quantum spins, their states, and associated observables. A natural segue led us to explore the Pauli matrices, which serve as key observables for spin-1/2 systems. The significance of this subject matter is multifold; not only do spins constitute the basis for spin-based qubits, a prominent category of qubit implementations,

but their study also fortifies our comprehension of the first three Postulates of Quantum Mechanics—namely, the state postulate, the observable postulate, and the measurement postulate.

Summary of Key Points

Our investigation started with an exploration of angular momentum, drawing distinctions between its classical and quantum characterizations. Through the resolution of eigenvalue equations for the Pauli matrices, X , Y , and Z , we were able to identify the eigenstates and eigenvalues, thereby building a robust mathematical framework for spin-1/2 systems.

Further, we engaged in a thorough discussion on general spin states and observables that can be aligned in any spatial direction. The geometric representation of these states was made transparent through the Bloch sphere, a construct vital for understanding the geometry of quantum states. The chapter closed with an examination of spin measurement techniques, from the pioneering Stern-Gerlach experiment to its modern-day variants, emphasizing their crucial role in the domain of quantum computing.

Implications and Applications

This chapter's insights are not purely academic; they have wide-ranging practical implications, particularly in the burgeoning field of quantum computing. Techniques like Larmor precession and Rabi oscillations serve as the backbone of stable and scalable quantum circuits. Moreover, the principles elucidated herein have applicability that extends into additional sectors such as quantum cryptography and quantum communication.

Closing Remarks

Mastering the quantum behavior of spins provides us with a refined vocabulary for describing microscopic phenomena. This enriched understanding not only challenges traditional classical viewpoints but is also instrumental in the rapidly advancing realm of quantum technologies. In upcoming chapters, we aim to construct a more general theoretical framework that encompasses the descriptions and dynamics of qubits, laying the groundwork for deeper inquiries into the subject.

Problem Set 2

2.1 Given the spin observable defined in Eq. 2.10:

$$\sigma_u = \begin{bmatrix} \cos \theta & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta \end{bmatrix},$$

calculate the following expected values:

(a) $\langle 0 | \sigma_u | 0 \rangle$,

(b) $\langle + | \sigma_u | + \rangle$ where $|+ \rangle = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle)$, and

(c) $\langle +_i | \sigma_u | +_i \rangle$ where $|+_i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$.

2.2 Given that σ_u is the spin observable defined in Eq. 2.10, and $|+u\rangle$ and $|-_u\rangle$ (Eq. 2.12) are its eigenvectors, using $|+u\rangle$ and $|-_u\rangle$ in as columns we can construct the following matrix:

$$U = \begin{bmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\phi} & \cos \frac{\theta}{2} e^{i\phi} \end{bmatrix}.$$

(a) Verify that U is unitary, i.e., $U^\dagger U = UU^\dagger = I$.

(b) Verify that $U^\dagger \sigma_u U = Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

The exercise illuminates the construction of a unitary transformation U that rotates a spin (or qubit) in the direction (θ, ϕ) to the z -direction.

2.3 A generic two-level quantum state can be expressed as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle,$$

where α and β are complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$.

Find θ and ϕ such that the above $|\psi\rangle$ is equivalent to $|+u\rangle$ in Eq. 2.12.

2.4 For the spin state $|+u\rangle$ (see Eq. 2.12) in the direction (θ, ϕ) , find the state vector of $|+u\rangle$ after the following rotations:

- (a) around the x -axis by π ,
- (b) around the y -axis by π ,
- (c) around the z -axis by π ,
- (d) around the x -axis by an angle γ ,
- (e) around the y -axis by an angle γ ,
- (f) around the z -axis by an angle γ ,
- (g) by π around the axis specified by polar angle α and azimuthal angle β , and
- (h) by γ around the axis specified by polar angle α and azimuthal angle β .

All rotations are assumed to be counterclockwise. You may use the Bloch sphere as a visualization tool.

2.5 Find a general formula for the square root of the 2×2 identity matrix I .

3. A Framework for Qubits and Qudits

Contents

3.1	Physical Qubit Systems	57
3.2	Qubit and Qudit States	57
3.2.1	General Qubit States	59
3.2.2	The Orthogonal Partner	59
3.2.3	The Bloch Sphere	60
3.2.4	* Qudit States	61
3.3	Change of Basis	62
3.3.1	Changing Between $ 0\rangle$ and $ 1\rangle$ to $ +\rangle$ and $ -\rangle$ Bases	63
3.3.2	* Change of Basis: Qudits	65
3.3.3	* Change from Computational Basis to Eigenvector Basis	67
3.4	* General Formulation of Quantum Measurement	68
3.4.1	Review of Basic Principles on Quantum Measurements	69
3.4.2	* A General Measurement Framework	69
3.4.3	* Projective Measurements for Observables	71
3.4.4	* Basis-Dependent Measurements	72
3.4.5	* Measuring in Alternative Bases	73
3.4.6	* Sampling Errors	74
3.4.7	* Advanced Topics in Quantum Measurements	76
3.5	* Application to Quantum State Tomography	77
3.5.1	* Extracting Parameters from Alice's Measurements	77
3.5.2	* Extracting Parameters from Bob's Measurements	77
3.5.3	* Reconstructing the Quantum State	78
3.6	Summary and Conclusions	78
	Problem Set 3	80

The chapter aims to develop a comprehensive framework for qubits, which serve as the cornerstone of quantum computing and quantum information systems. While the preceding chapters introduced key concepts using specific examples such as photons and electron spins, this chapter generalizes these ideas to accommodate all

forms of qubits. We also extend our discussion to include qudits, or general d -level quantum systems, where relevant.

Two focal topics—change of basis and quantum measurement theory—will be explored in depth. Quantum state tomography will be highlighted as a real-world application to demonstrate these principles.

This framework serves as an essential link, facilitating a smooth transition from fundamental concepts to more advanced topics in the quantum realm.

3.1 Physical Qubit Systems

In the preceding chapters, we discussed photon polarization and $\frac{1}{2}$ spins as forms of physical qubits, which are two-level quantum systems. However, there are several other platforms being actively explored for qubit implementation in quantum computing. Some of these are intrinsically multilevel systems but offer two states that can be effectively isolated, thereby enabling their use as qubits.

The following table (Table 3.1) provides a list of qubit platforms, along with the method of information encoding and the basis states relevant to each system. This table aims to give an overview of the breadth of qubit platforms currently under investigation, thus showcasing the versatility and adaptability of qubit systems.

These qubit implementations are currently subjects of intense research and development, each with its own set of advantages and limitations [52]. Factors such as noise rates, controllability, scalability, and cost vary significantly among different qubit platforms. While some may offer excellent controllability, others might excel in terms of low noise or cost-effectiveness. As the field of quantum computing continues to mature, it remains an open question which of these qubit implementations will emerge as the most practical and widely adopted for large-scale quantum information processing.

Exercise 3.1 Conduct an investigation into the qubit systems enumerated in Table 3.1 as well as others you come across. Your investigation should focus on the following aspects for each qubit system:

- Current state-of-the-art developments
- Advantages and disadvantages, including noise rates, controllability, scalability, and cost
- Leading companies and research groups actively working on these platforms

Compile your findings into concise summaries for each qubit system to gain a comprehensive understanding of the current landscape of quantum computing.

3.2 Qubit and Qudit States

Despite the various physical manifestations of qubits, they can all be characterized by a shared mathematical model grounded in quantum mechanics.

Qubit System	Information Encoding	Basis States
Photon polarization	Polarization states	Vertical/horizontal polarizations, or left/right circular polarizations
Photon path	Dual-rail encoding	Path A and path B
Squeezed light	Quadrature states	Amplitude- and phase-squeezed states
Superconductor charge qubit	Charge	Uncharged and charged islands
Superconductor flux qubit	Current	Clockwise and counterclockwise current
Superconductor transmon qubit	Energy levels, with large capacitance	Ground and excited states
Superconductor fluxonium qubit	Energy levels, with large inductance	Ground and excited states
Trapped ion	Electron internal states, hyperfine levels	Ground and excited states
Neutral atoms, Rydberg atoms	Hyperfine levels, Rydberg states	Ground and excited states
Molecular qubit	Molecular states	Rotational or vibrational states
Quantum dot - spin	Electron spin	Up/down spin states
Quantum dot pair	Electron localization	Left dot and right dot
Nitrogen-vacancy (N-V) center	Electron spin	Electron spin projection ($m_s = 0$ or -1)
Cat qubit	Superpositions of coherent states	Coherent state superpositions
Topological system (anyons)	Braiding of excitations	Fusion basis states

Table 3.1: Examples of Qubit Implementations

3.2.1 General Qubit States

The State Postulate posits that the state of a quantum system is encapsulated by a unit vector within a Hilbert space. For qubits, this space constitutes a two-dimensional complex vector space, denoted as \mathbb{C}^2 .

In the realm of quantum computing, the basis states for an individual qubit are typically represented as $|0\rangle$ and $|1\rangle$, reminiscent of the classical binary states 0 and 1. A general qubit state can be depicted as a linear combination of these basis states:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \quad (3.1)$$

In this expression, α and β are complex coefficients. Due to the necessity for the state vector to be normalized, we have the constraint $|\alpha|^2 + |\beta|^2 = 1$. This arises from $|\alpha|^2$ and $|\beta|^2$ representing the probabilities of measuring the qubit in states $|0\rangle$ and $|1\rangle$, respectively, and their cumulative sum equating to 1.

1 Global Phase and Relative Phase

The phase of a complex number, represented as $re^{i\phi}$, is indicated by the ϕ value. Introducing a global phase to a quantum state applies an identical phase factor to both basis states, which leaves the measurement probabilities of each state unaltered. Therefore, a state such as:

$$|\psi'\rangle = \alpha e^{i\phi}|0\rangle + \beta e^{i\phi}|1\rangle,$$

is congruent to $|\psi\rangle$, since both $|\alpha e^{i\phi}|^2$ and $|\beta e^{i\phi}|^2$ equate to their non-phased counterparts. Consequently, global phase lacks physical significance. However, relative phase is of importance as it can influence measurement outcomes across varied bases.

2 Expressing Qubit States with Real Numbers

Acknowledging that global phase is non-consequential and that the state vector is normalized, it is feasible to convey a general qubit state using a mere two real parameters. Adopting the notations $\alpha = \cos \frac{\theta}{2}$ and $\beta = \sin \frac{\theta}{2} e^{i\phi}$, we obtain:

$$|\psi\rangle = \cos \frac{\theta}{2}|0\rangle + \sin \frac{\theta}{2} e^{i\phi}|1\rangle. \quad (3.2)$$

With this parameterization, both the normalization condition and the relative phase factor are inherently satisfied.

Equation 3.2 mirrors $|+_u\rangle$ from Eq. 2.12, illustrating a $\frac{1}{2}$ -spin oriented in the direction (θ, ϕ) . Now this has been derived devoid of any explicit reliance on the intrinsic nature of the qubits.

Exercise 3.2 Show that the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ represents a $\frac{1}{2}$ -spin oriented in the direction: $\hat{u} = (2\operatorname{Re}(\alpha\beta^*), 2\operatorname{Im}(\alpha\beta^*), |\alpha|^2 - |\beta|^2)$. (3.3)

3.2.2 The Orthogonal Partner

The state which stands orthonormal to $|\psi\rangle$ is:

$$|\psi_{\perp}\rangle = \beta^* |0\rangle - \alpha^* |1\rangle, \quad (3.4)$$

ensuring both $\langle\psi_{\perp}|\psi_{\perp}\rangle = 1$ and $\langle\psi|\psi_{\perp}\rangle = 0$. Here β^* denotes the complex conjugate of β . It is worth noting that the analogous state $e^{i\phi} |\psi_{\perp}\rangle$, especially $-|\psi_{\perp}\rangle$, is also orthonormal to $|\psi\rangle$. In its (θ, ϕ) representation, we have:

$$|\psi_{\perp}\rangle = -\sin \frac{\theta}{2} |0\rangle + \cos \frac{\theta}{2} e^{i\phi} |1\rangle, \quad (3.5)$$

which corresponds to $|-_u\rangle$ in Eq. 2.12.

Exercise 3.3 Given the states:

$$|\psi_1\rangle = |0\rangle,$$

$$|\psi_2\rangle = -\frac{1}{2} |0\rangle + \frac{\sqrt{3}}{2} |1\rangle,$$

$$|\psi_3\rangle = -\frac{1}{2} |0\rangle - \frac{\sqrt{3}}{2} |1\rangle,$$

prove that:

$$|\langle\psi_1|\psi_2\rangle|^2 = |\langle\psi_2|\psi_3\rangle|^2 = |\langle\psi_3|\psi_1\rangle|^2 = \frac{1}{4}.$$

3.2.3 The Bloch Sphere

In § 2.4, we introduced the Bloch sphere in the context of electron spin. Beyond spin, the Bloch sphere provides a versatile visualization for general qubit states. Each point on this sphere represents a distinct qubit state, denoted as $|\psi\rangle$. The vector emanating from the origin to this point, symbolized as \hat{u} , is the Bloch vector. This vector is defined by two angles: the polar angle θ and the azimuthal angle ϕ . Likewise, the state vector $|\psi\rangle$ is parameterized using these same angles. Fig. 3.1 elucidates these definitions alongside highlighting key points on the sphere.

The expectation values of the Pauli spin operators X , Y , and Z in the state $|\psi\rangle$ (refer to Eq. 3.2) are given by:

$$\langle X \rangle = \sin \theta \cos \phi, \quad (3.6a)$$

$$\langle Y \rangle = \sin \theta \sin \phi, \quad (3.6b)$$

$$\langle Z \rangle = \cos \theta. \quad (3.6c)$$

These coincide with the components of \hat{u} , allowing $|\psi\rangle$ to be interpreted as representing the angular momentum of the spin in three dimensions.

Exercise 3.4 Prove Eq. 3.6.

An important detail to note is the presence of the term $\frac{\theta}{2}$ within the formulation of $|\psi\rangle$. Moreover, it's intriguing to observe that orthogonal states, such as $|+\rangle$ and $|-\rangle$, occupy antipodal positions on the sphere. This implies they are separated by an angle of 180° , contrary to the 90° separation that might be more intuitively

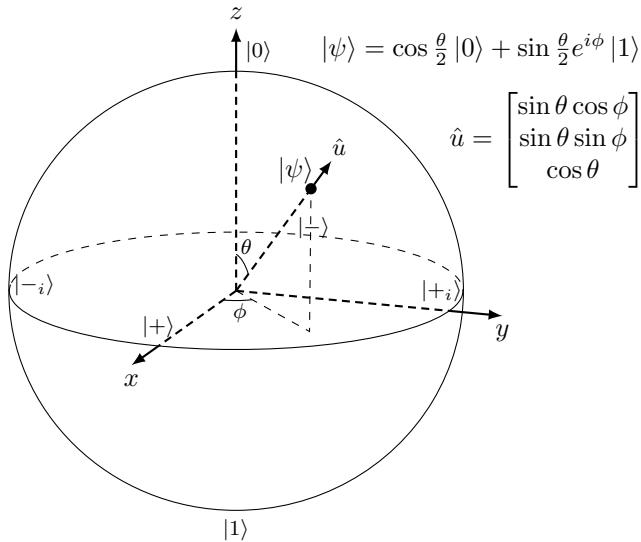


Figure 3.1: Visualization of qubit states on the Bloch Sphere

anticipated. This particular aspect is related to the notion of $\frac{1}{2}$ -spin representing an angular momentum.

3.2.4 *Qudit States

Qutrits and qudits denote quantum systems with more than two levels. While qubits are ubiquitous in quantum computing, qutrits (three-level systems) and qudits (general d -level systems) offer more intricate and potentially versatile quantum states. Various quantum systems, including superconducting qubits, Rydberg atoms, and NV centers, possess more than two inherent states. However, control, calibration, and maintaining coherence of these states become substantially more complex and demanding.

Despite these challenges, there's growing enthusiasm in the research community to explore qudits as they present potential pathways for advanced quantum computational techniques. As the study progresses and whenever it benefits pedagogical discussions, the standard qubit framework will be broadened to include discussions on qudits.

In addition, an n -qubit system is mathematically equivalent to a $d = 2^n$ qudit. This equivalence means that formulations developed for qudits can also be applied to multiple qubits that are combined and treated as a single quantum system, broadening the applicability and relevance of qudit-based theoretical frameworks.

The general state of a qudit can be expressed as:

$$|\psi\rangle = \sum_{j=0}^{d-1} c_j |j\rangle, \quad (3.7)$$

where $c_j \in \mathbb{C}$. This state's normalization constraint is given by:

$$\langle \psi | \psi \rangle = \sum_{j=0}^{d-1} |c_j|^2 = 1. \quad (3.8)$$

The basis states' orthogonality ensures:

$$\langle j | k \rangle = \delta_{j,k}, \quad (3.9)$$

with $\delta_{j,k}$ being the Kronecker delta function.

The completeness condition for the basis states is expressed as:

$$\sum_{j=0}^{d-1} |j\rangle \langle j| = I. \quad (3.10)$$

A quantum state in d dimensions can be parameterized by $2(d-1)$ real parameters due to the global phase factor and the normalization condition. Consequently, the state representation becomes:

$$|\psi\rangle = \begin{bmatrix} \cos \theta_1 \\ \cos \theta_2 \sin \theta_1 e^{i\phi_1} \\ \cos \theta_3 \sin \theta_1 \sin \theta_2 e^{i\phi_2} \\ \vdots \\ \cos \theta_{d-1} \sin \theta_1 \cdots \sin \theta_{d-2} e^{i\phi_{d-2}} \\ \sin \theta_1 \cdots \sin \theta_{d-2} \sin \theta_{d-1} e^{i\phi_{d-1}} \end{bmatrix}. \quad (3.11)$$

Exercise 3.5 * For the state $|\psi\rangle$ in Eq. 3.11, identify at least one vector from the set of $d - 1$ vectors that are orthonormal to it.

3.3 Change of Basis

Understanding qubit states and operators across different bases is fundamental to quantum computing. The importance of bases can be distilled into two primary considerations:

1. **Perspectives for Quantum States:** Alternate bases often illuminate novel perspectives on quantum states, offering insights pivotal for specific quantum operations or algorithms. In this context, while the underlying physical quantum states and observables remain constant, their vector or matrix representations undergo transformation. This idea can be likened to the classical concept of reference frames in Newtonian physics. Inspecting quantum states from varied “viewpoints” mirrors the act of changing the basis for the state vectors.
2. **Quantum State Evolution as Basis Transformation:** The operations of quantum gates or other unitary transformations on quantum states can be mathematically interpreted as a change of basis in the respective vector spaces. In quantum computing, we often work in the computational basis, but the quantum state itself exists in a Hilbert space that can be represented in any orthonormal basis. Unlike the previous scenario, here the physical quantum states undergo genuine changes.

The rudiments of basis change were introduced in § 1.4.5, utilizing photons as exemplars. This section will expand upon those foundational concepts, aiming to furnish a rigorous framework for changes in basis. In § 4.1, we will investigate the unitary time-evolution that can be viewed as a change of basis.

3.3.1 Changing Between $|0\rangle$ and $|1\rangle$ to $|+\rangle$ and $|-\rangle$ Bases

To build intuition, let's examine a straightforward scenario.

1 Quantum States in Two Bases

A single qubit's quantum states are two-dimensional, necessitating a basis of two orthogonal, normalized states. The standard computational basis utilizes states $|0\rangle$ and $|1\rangle$. Meanwhile, the $\{|+\rangle, |-\rangle\}$ basis (often referred to as the Hadamard basis) defines them as:

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \quad (3.12)$$

A general qubit state in the computational basis, $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, can be expressed in the $\{|+\rangle, |-\rangle\}$ basis as:

$$|\psi\rangle = \frac{\alpha + \beta}{\sqrt{2}}|+\rangle + \frac{\alpha - \beta}{\sqrt{2}}|-\rangle. \quad (3.13)$$

This provides new coefficients for the states in the revised basis.

Exercise 3.6 Express the θ, ϕ parametrized $|\psi\rangle$ given in Eq. 3.2 in the $|+\rangle, |-\rangle$ basis.

2 Basis Transformation Operator

Transitioning from one orthonormal basis to another in quantum mechanics corresponds with a unitary operator, as unitary transformations preserve the inner product and hence the orthonormality of the basis. (See Appendix C for the definition and properties of unitary operators.) The basis transformation operator U for transitioning from $\{|0\rangle, |1\rangle\}$ to $\{|+\rangle, |-\rangle\}$ corresponds to a quantum evolution where $|0\rangle$ transforms to $|+\rangle$ and $|1\rangle$ to $|-\rangle$. This operator takes a linear combination of states $|0\rangle$ and $|1\rangle$ to the corresponding combination of $|+\rangle$ and $|-\rangle$. Hence,

$$|+\rangle = U|0\rangle, \quad |-\rangle = U|1\rangle. \quad (3.14)$$

Consequently, the operator U is:

$$U = |+\rangle\langle 0| + |-\rangle\langle 1|. \quad (3.15)$$

Proof. The definition of U provides:

$$|+\rangle = U|0\rangle \Rightarrow |+\rangle\langle 0| = U|0\rangle\langle 0|, \quad (3.16a)$$

$$|-\rangle = U|1\rangle \Rightarrow |-\rangle\langle 1| = U|1\rangle\langle 1|. \quad (3.16b)$$

Combining the two equations gives:

$$|+\rangle\langle 0| + |-\rangle\langle 1| = U(|0\rangle\langle 0| + |1\rangle\langle 1|). \quad (3.17)$$

Utilizing the completeness property $|0\rangle\langle 0| + |1\rangle\langle 1| = I$, we deduce Eq. 3.15. \square

Expanding Eq. 3.15, we obtain another expression for U that directly mirrors its matrix form:

$$U = \frac{1}{\sqrt{2}}(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| - |1\rangle\langle 1|). \quad (3.18)$$

Exercise 3.7 Validate that $U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ performs the transformation in Eq. 3.14.

3 Inverse Transformation

Applying U^{-1} to $|+\rangle, |-\rangle$ brings them back to $|0\rangle, |1\rangle$. Given that U is unitary (meaning $U^{-1} = U^\dagger$):

$$|0\rangle = U^\dagger |+\rangle, \quad |1\rangle = U^\dagger |-\rangle. \quad (3.19)$$

Exercise 3.8 Verify that U in Eq. 3.15 is unitary using the bra-ket notation.

Hint: $U^\dagger = |0\rangle\langle +| + |1\rangle\langle -|$.

4 Basis Rotation

As depicted in Fig. 3.2, a basis change can be visualized as a rotation within the Hilbert space's coordinate system. While the vector $|\psi\rangle$ stays physically unchanged, its components with respect to the new basis vary. This variation effectively presents the same quantum state but in a different mathematical form.

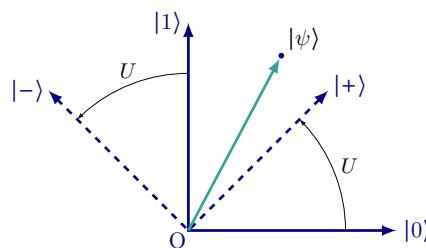


Figure 3.2: Basis Rotation $\{|0\rangle, |1\rangle\} \rightarrow \{|+\rangle, |-\rangle\}$

Similar to how a geometric rotation preserves vector length and relative angles in the coordinate system, a basis change via a unitary transformation keeps the inner product intact. This preservation is the reason a change of basis is often conceptualized as a basis rotation.

5 Transformation of Vector Components

The vector components can be transformed directly using U^\dagger :

$$\begin{bmatrix} \frac{\alpha+\beta}{\sqrt{2}} \\ \frac{\alpha-\beta}{\sqrt{2}} \end{bmatrix} = U^\dagger \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (3.20)$$

Exercise 3.9 For a given quantum state $|\psi\rangle = \frac{1}{2}|+\rangle + i\frac{\sqrt{3}}{2}|-\rangle$, compute its representation in $\{|0\rangle, |1\rangle\}$ basis.

6 Computing Vector Components through Projection

In addition to the matrix representation of the transformation operator U and its application to state vectors, there is an intrinsic method for calculating the amplitudes of a quantum state in a specific basis. This approach involves decomposing (or projecting) the state $|\psi\rangle$ into an orthonormal basis. For the standard basis $\{|0\rangle, |1\rangle\}$, the state can be expressed as

$$|\psi\rangle = \langle 0|\psi\rangle |0\rangle + \langle 1|\psi\rangle |1\rangle, \quad (3.21)$$

where the components of $|\psi\rangle$ are the projections $\langle 0|\psi\rangle$ and $\langle 1|\psi\rangle$. With $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, these yield α and β as expected.

Similarly, for the basis $\{|+\rangle, |-\rangle\}$, the state can be expressed as

$$|\psi\rangle = \langle +|\psi\rangle |+\rangle + \langle -|\psi\rangle |-\rangle, \quad (3.22)$$

$$\langle +|\psi\rangle = \frac{\alpha + \beta}{\sqrt{2}}, \quad \langle -|\psi\rangle = \frac{\alpha - \beta}{\sqrt{2}}. \quad (3.23)$$

3.3.2 *Change of Basis: Qudits

After acquainting ourselves with the fundamental principles and operations of basis change using the example $\{|0\rangle, |1\rangle\} \rightarrow \{|+\rangle, |-\rangle\}$, we now extend them to general qudit systems, where the index (e.g., i or j below) extends to $\{0, 1, \dots, d-1\}$.

1 Definition and Operator

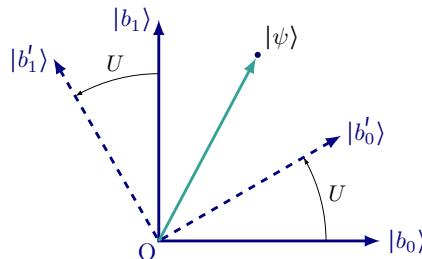


Figure 3.3: Illustration of Basis Rotation $\{b_i\} \rightarrow \{b'_i\}$

Given two complete and orthonormal bases $\{|b_i\rangle\}$ and $\{|b'_i\rangle\}$, their orthogonality and completeness relations are:

$$\langle b_i | b_j \rangle = \langle b'_i | b'_j \rangle = \delta_{ij}, \quad \sum_i |b_i\rangle \langle b_i| = \sum_i |b'_i\rangle \langle b'_i| = I. \quad (3.24)$$

To change from the basis $\{|b_i\rangle\}$ to $\{|b'_i\rangle\}$, a unitary operator U is defined such that $|b'_i\rangle = U|b_i\rangle$. Specifically, U is:

$$U = \sum_i |b'_i\rangle \langle b_i| = \sum_{i,j} \langle b_j | b'_i \rangle |b_j\rangle \langle b_i|. \quad (3.25)$$

Exercise 3.10 Derive the formulas of U in Eq. 3.25 from its definition $|b'_i\rangle = U|b_i\rangle$.

Hint: see the proof for the $\{|0\rangle, |1\rangle\} \rightarrow \{|+\rangle, |-\rangle\}$ case.

2 Vectors and Matrices under Change of Basis

While physical properties, such as quantum states and observables, remain invariant across distinct bases, their vector or matrix representations are transformed by U upon a basis change.

A general qudit state $|\psi\rangle$ can be expressed in either of the bases:

$$|\psi\rangle = \sum_i c_i |b_i\rangle = \sum_i c'_i |b'_i\rangle. \quad (3.26)$$

The coefficients or vector components of $|\psi\rangle$ are obtained as:

$$c_i = \langle b_i | \psi \rangle, \quad c'_i = \langle b'_i | \psi \rangle. \quad (3.27)$$

Considering $\{c_i\}$ as a column vector $|c\rangle$ and $\{c'_i\}$ as $|c'\rangle$, we have:

$$|c\rangle = U|c'\rangle, \quad |c'\rangle = U^\dagger|c\rangle. \quad (3.28)$$

Inner products are scalars and, therefore, remain constant across different bases:

$$\langle u' | v' \rangle = \langle u | v \rangle, \quad (3.29)$$

given that $|u'\rangle = U^\dagger|u\rangle$ and $|v'\rangle = U^\dagger|v\rangle$ according to Eq. 3.28.

Exercise 3.11 Prove the invariance of the inner product under a unitary transformation (see Eq. 3.29).

For an observable A , $\langle u | A | u \rangle$ represents the expected value (statistical average) of A in the state $|u\rangle$ (see § 1.5.3). It is also a scalar and, therefore, should remain constant across different bases. This requires the matrix form of A to transform according to:

$$A' = U^\dagger A U. \quad (3.30)$$

With this transformation, $\langle u|A|u \rangle$ is indeed invariant under a unitary transformation:

$$\langle u'|A'|u' \rangle = \langle U^\dagger u|U^\dagger AU|U^\dagger u \rangle \quad (3.31a)$$

$$= \langle u|UU^\dagger AUU^\dagger|u \rangle \quad (3.31b)$$

$$= \langle u|A|u \rangle. \quad (3.31c)$$

In § 3.4.4, we will demonstrate the basis independence of $\langle u|A|u \rangle$ through an example.

Exercise 3.12 Demonstrate that, if two observables commute in the original basis ($[A, B] = 0$), they will also commute in the transformed basis ($[UAU^\dagger, UBU^\dagger] = 0$).

3.3.3 *Change from Computational Basis to Eigenvector Basis

The action of transitioning to the eigenvector basis of a Hermitian operator is termed as “diagonalizing the operator”. This nomenclature stems from the fact that, in the eigenvector basis, the operator is represented as a diagonal matrix with its eigenvalues on the diagonal. Such a transformation finds significant applications in quantum computing. Prominent among these are the quantum phase estimation and Hamiltonian simulation algorithms. Furthermore, as discussed in § 3.4.5, measurements in quantum computing often employ such basis transformations.

Given H is a Hermitian operator with eigenvalues $\{\lambda_i\}$ and normalized eigenvectors $\{|\phi_i\rangle\}$:

$$H |\phi_i\rangle = \lambda_i |\phi_i\rangle, \quad (3.32)$$

$\{|\phi_i\rangle\}$ form a complete, orthonormal basis.

To change basis from the computational basis $\{|i\rangle\}$ to the H eigenbasis $\{|\phi_i\rangle\}$, we can use the unitary basis transformation operator U given by:

$$|\phi_i\rangle = U |i\rangle, \quad U = \sum_i |\phi_i\rangle \langle i|. \quad (3.33)$$

As a consequence of the spectral decomposition theorem, the matrix form of H is diagonal in the $\{|\phi_i\rangle\}$ basis:

$$H = \sum_i \lambda_i |\phi_i\rangle \langle \phi_i|, \quad (3.34)$$

while the matrix form of $U^\dagger H U$ is diagonal in the computational basis:

$$U^\dagger H U = \sum_i \lambda_i |i\rangle \langle i|. \quad (3.35)$$

Eigenbases for Pauli Operators

The eigenbases of the Pauli operators play an important role in quantum computations and measurements.

- The eigenbasis for Pauli Z is the computational basis itself, consisting of the eigenvectors $|0\rangle$ and $|1\rangle$.
- The eigenbasis for Pauli X consists of the states $|+\rangle$ and $|-\rangle$. The unitary transformation relating this basis to the computational basis is the Hadamard operator H , defined as

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (3.36)$$

It can be verified that $|+\rangle = H|0\rangle$ and $|-\rangle = H|1\rangle$. In addition, the diagonalization relationship $Z = HXH$ holds. This basis is sometimes referred to as the X basis or the Hadamard basis.

Note that here, H corresponds to U from Eq. 3.33, and X corresponds to H from Eq. 3.32. We have repurposed H as it is the conventional symbol for the Hadamard operator.

- The eigenvectors of the Pauli Y operator are $|+_i\rangle$ and $|-_i\rangle$. For the eigenbasis, we often use the equivalent vectors $|+_i\rangle$ and $i|-_i\rangle$ for the convenience of calculation. The unitary operator that transforms from this eigenbasis to the computational basis is $R_x(-\frac{\pi}{2})$, also denoted simply as R_x . This operator is defined as (see § 5.2.5):

$$R_x = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}. \quad (3.37)$$

One can verify that $|+_i\rangle = R_x|0\rangle$ and $i|-_i\rangle = R_x|1\rangle$. Furthermore, the diagonalization relationship $Z = R_x^\dagger Y R_x$ holds.

Exercise 3.13 Given the Hermitian operator:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},$$

- (a) Find its eigenvalues and the corresponding eigenvectors.
- (b) If a system is in a state $|\psi\rangle = |0\rangle$, express this state in the eigenbasis of A .
- (c) Consider measuring the state $|\psi\rangle = |0\rangle$ in the eigenbasis of A which is equivalent to measurement using A as observable. What are the probabilities associated with each eigenvalue?

3.4 * General Formulation of Quantum Measurement

Quantum measurement plays an essential role in quantum computing. It serves as the bridge that transfers information from the quantum realm to the classical world. Moreover, quantum measurement is an indispensable tool for investigating decoherence, a major source of error in contemporary quantum computing systems. Although intricate, mastering this topic is essential for understanding quantum

system operations. In this section, we will formulate a general quantum measurement framework based on basic principles of observable and measurement in quantum mechanics discussed in Chapters 1 and 2.

3.4.1 Review of Basic Principles on Quantum Measurements

The Observable Postulate and the Measurement Postulate are two fundamental postulates in quantum mechanics that describe the behavior of quantum systems and the act of measuring those systems.

The Observable Postulate states that every observable quantity in quantum mechanics is represented by a Hermitian operator. An observable quantity is a physical property of a quantum system that can be measured, such as position, momentum, energy, or spin. The Hermitian operator associated with an observable quantity is called an observable operator, and its eigenvalues represent the possible outcomes of a measurement of that observable quantity.

The Measurement Postulate, on the other hand, describes the process of measuring an observable quantity of a quantum system. According to the Composition Postulate, the act of measuring an observable quantity of a quantum system will cause the system to collapse into one of the eigenstates of the observable operator associated with that observable quantity, and the corresponding eigenvalue will be the result of the measurement. This collapse is also known as the wavefunction collapse, and it is a fundamental feature of quantum mechanics.

These postulates lay the groundwork for understanding quantum measurement in its most straightforward form. However, modern quantum computing and quantum information often employ more advanced measurement schemes such as POVM (Positive Operator-Valued Measure), weak measurements, and nondestructive measurements. In order to accommodate such complex scenarios, it is necessary to extend these standard postulates into a general quantum measurement framework.

3.4.2 *A General Measurement Framework

Quantum measurements are typically characterized by a set of measurement operators $\{M_i\}$, representing potential measurement outcomes.

1 Measurement Probability

Given a quantum system in state $|\psi\rangle$, the probability of realizing outcome i upon measurement is:

$$P_i = \langle \psi | M_i^\dagger M_i | \psi \rangle. \quad (3.38)$$

This formula mirrors the squared magnitude of the vector $M_i |\psi\rangle$, denoted as $\|M_i |\psi\rangle\|^2$.

The product $M_i^\dagger M_i$ is positive semidefinite, that is, a Hermitian matrix with non-negative eigenvalues, ensuring all probabilities $P_i \geq 0$.

Exercise 3.14 Demonstrate that a matrix defined as $M_i^\dagger M_i$ is positive semidefinite.

For the total probability to sum up to 1, i.e., $\sum P_i = 1$, the operators $\{M_i\}$ must adhere to the completeness condition:

$$\sum_i M_i^\dagger M_i = I. \quad (3.39)$$

Exercise 3.15 Show that when M_i meets the completeness requirement Eq. 3.39, the sum $\sum P_i$ equals 1.



In this broad portrayal of quantum measurement, the operators $\{M_i\}$ need not be orthogonal.

2 Measurement Average

If a numerical value λ_i corresponds to each M_i , the expected value of λ is:

$$\langle \lambda \rangle = \sum_i P_i \lambda_i = \sum_i \langle \psi | M_i^\dagger M_i | \psi \rangle \lambda_i. \quad (3.40)$$

3 Post Measurement State

A salient feature of quantum measurement is the alteration of the quantum state upon measurement. After the measurement, the system transitions to:

$$|\psi_i\rangle = \frac{M_i |\psi\rangle}{\sqrt{P_i}}. \quad (3.41)$$

The factor $\sqrt{P_i}$ in the denominator guarantees the normalization of $|\psi_i\rangle$, ensuring it remains a legitimate quantum state.

4 Measurement Probability via State Overlap

The likelihood of observing a specific measurement result can be deduced from the overlap (inner product) of the pre- and post-measurement states:

$$P_i = |\langle \psi_i | \psi \rangle|^2. \quad (3.42)$$

Exercise 3.16 Using Eq. 3.41, confirm that the probability of measuring $|\psi\rangle$ and subsequently obtaining $|\psi_i\rangle$ is given by Eq. 3.42.

5 Non-Unitary Transformation

It is pertinent to point out that Eq. 3.41 does not generally represent a unitary transformation, underscoring a distinctive trait of quantum measurement. Should we perform back-to-back measurements with the identical M_i on the post-measurement state (assuming no intermediate state alterations, excluding general POVM measurements), outcomes will be consistent, and the state remains unperturbed.

6 POVM Measurements

The measurement framework outlined here can be extended to the Positive Operator-Valued Measure (POVM) formalism. In the POVM context, the focus is on the measurement outcomes rather than the state of the system after the measurement has been performed. Thus, $\{M_i^\dagger M_i\}$ is replaced with a set of POVM operators $\{E_i\}$, each being a positive semidefinite Hermitian operator that corresponds to a potential measurement outcome, with $\sum_i E_i = I$. Note that given E_i , its decomposition $E_i = M_i^\dagger M_i$ is non-unique. All the preceding equations in this section are still valid except Eq. 3.41.

3.4.3 *Projective Measurements for Observables

In light of our general formulation, measurements pertaining to a quantum observable emerge as a special case known as projective measurement.

Consider an observable H , represented by a Hermitian operator. This operator possesses a collection of real eigenvalues λ_i with corresponding eigenvectors $|\phi_i\rangle$. These eigenvectors construct a complete orthonormal basis. Accordingly, the measurement operators M_i become the projection operators associated with the basis states $|\phi_i\rangle$:

$$M_i = \Pi_i = |\phi_i\rangle\langle\phi_i|. \quad (3.43)$$

Being projection operators, beyond the completeness relationship in Eq. 3.39, the set $\{M_i\}$ adheres to the orthogonal and idempotent properties:

$$M_i M_j = \delta_{ij} M_i. \quad (3.44)$$

Exercise 3.17 Validate Eq. 3.44.

Exercise 3.18 Employing the relationship $\sum_i |\phi_i\rangle\langle\phi_i| = I$, verify that Eq. 3.39 is valid for projective measurements.

Subsequently, the probability formula in Eq. 3.38 for acquiring outcome λ_i is recast as:

$$P_i = \langle\psi|\Pi_i|\psi\rangle = |\langle\psi|\phi_i\rangle|^2. \quad (3.45)$$

Likewise, the formula for the expected value of M in Eq. 3.40 transforms to:

$$\langle H \rangle = \sum_i P_i \lambda_i = \langle\psi|H|\psi\rangle, \quad (3.46)$$

making use of the spectral decomposition property $H = \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|$.

Exercise 3.19 Confirm that Eq. 3.40 yields Eqs. 3.45 and 3.46.

Finally, the post-measurement state from Eq. 3.41 simplifies to $|\phi_i\rangle$. In this sense, the projection operator essentially selects the respective basis state, i.e., projects the original state onto the basis state.

Our discussions highlight that when the measurement operators $\{M_i\}$ are projection operators linked to the eigenstates of an observable, the overarching measurement framework naturally aligns with observable-centric measurement.

3.4.4 * Basis-Dependent Measurements

Basis dependence is an essential facet of quantum measurement. When quantum systems are measured in different bases, their outcomes differ based on the chosen basis.

To elucidate, a quantum state measured in a specific basis is essentially being measured against that basis's states. For instance, when measuring the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ in the computational basis, it splits into post-measurement states $|0\rangle$ and $|1\rangle$, with associated probabilities of $|\alpha|^2$ and $|\beta|^2$. This phenomenon is depicted in Fig. 3.4, part (a).

In a similar vein, if $|\psi\rangle$ is measured in the $\{|+\rangle, |-\rangle\}$ basis, as illustrated in part (b), it results in post-measurement states $|+\rangle$ and $|-\rangle$ with probabilities $\frac{|\alpha+\beta|^2}{2}$ and $\frac{|\alpha-\beta|^2}{2}$.

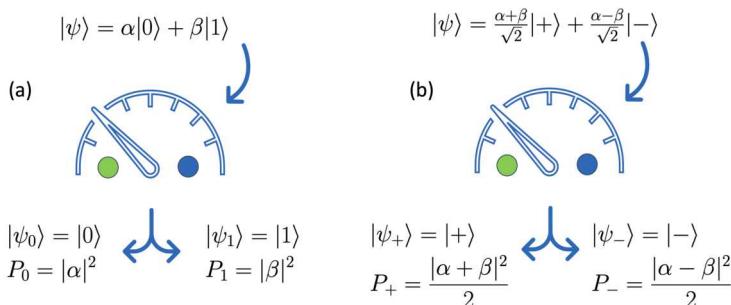


Figure 3.4: Measurements in $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$ Bases

Traditionally, distinct bases measurements correspond to different observables. For the purpose of quantum computing, let's assign $+1$ and -1 as standard measurement values. Thus, the pertinent observables can be expressed as:

$$M_{0,1} = |0\rangle\langle 0| - |1\rangle\langle 1| = Z, \text{ corresponding to the } \{|0\rangle, |1\rangle\} \text{ basis}; \quad (3.47a)$$

$$M_{+,-} = |+\rangle\langle +| - |-\rangle\langle -| = X, \text{ associated with the } \{|+\rangle, |-\rangle\} \text{ basis}. \quad (3.47b)$$

Given this convention, the measurements in the provided example can also be dubbed as the Z and X measurements.

When an observable M undergoes measurement in different bases, its expected value, $\langle M \rangle$, remains constant (see § 3.3.2). In the above example, the expected value of X calculated in the $\{|+\rangle, |-\rangle\}$ basis through probability argument is

$$\langle X \rangle = (+1)P_+ + (-1)P_- = \frac{|\alpha + \beta|^2}{2} - \frac{|\alpha - \beta|^2}{2} = \alpha\beta^* + \alpha^*\beta. \quad (3.48)$$

But as demonstrated in § 1.5.3, $\langle X \rangle$ is given by $\langle \psi | X | \psi \rangle$, which is independent of the measurement basis. Indeed, one can verify that $\langle \psi | X | \psi \rangle = \alpha\beta^* + \alpha^*\beta$ by computing the inner product directly in the computational basis.

Exercise 3.20 The expected value of Z calculated through a probability argument is:

$$\langle Z \rangle = (+1)P_0 + (-1)P_1 = |\alpha|^2 - |\beta|^2. \quad (3.49)$$

Verify $\langle \psi | Z | \psi \rangle$ yields the same value in both the computational and $\{|+\rangle, |-\rangle\}$ bases.

3.4.5 *Measuring in Alternative Bases

In quantum computing, the computational basis is the de facto standard for measurements, primarily due to the native support from most quantum hardware. Yet, certain scenarios call for measurements in non-computational bases. In such instances, unitary transformations, or basis rotations, are employed to shift to the desired basis before carrying out the measurement.

Suppose one wishes to measure a qudit state, denoted as $|\psi\rangle$, within an orthonormal basis $\{|\phi_i\rangle\}$. The initial task is to identify a suitable unitary transformation, represented by U , that facilitates the transition from the targeted measurement basis $\{|\phi_i\rangle\}$ to the computational basis $\{|i\rangle\}$, up to a global phase factor:

$$U = \sum_i |i\rangle\langle\phi_i|. \quad (3.50)$$

Upon establishing U , the subsequent step involves measuring the state $U|\psi\rangle$ within the computational basis. Interestingly, the outcome probabilities yielded from this process mirror those obtained by measuring $|\psi\rangle$ directly in the $\{|\phi_i\rangle\}$ basis. The congruence of the results can be verified as follows:

$$P_i = |\langle i|U|\psi\rangle|^2 \quad (3.51a)$$

$$= \left| \sum_j \langle i|j\rangle \langle\phi_j|\psi\rangle \right|^2 \quad (3.51b)$$

$$= |\langle\phi_i|\psi\rangle|^2. \quad (3.51c)$$

$$|\psi\rangle \xrightarrow{\text{ } \begin{array}{|c|} \hline \{\mid\phi_i\}\rangle \\ \hline \end{array} } P_i \Rightarrow |\psi\rangle \xrightarrow{\text{ } \begin{array}{|c|} \hline U \\ \hline \end{array} } \xrightarrow{\text{ } \begin{array}{|c|} \hline \{\mid i\}\rangle \\ \hline \end{array} } P_i$$

An essential distinction to highlight is that, post-measurement, the state achieved in the computational basis, $|i\rangle$, does not coincide with the anticipated state, $|\phi_i\rangle$, within the chosen measurement basis. Nevertheless, within the quantum computing context, this incongruence is typically inconsequential since the primary emphasis

is on obtaining the measurement readout. If a scenario necessitates the post-measurement state, employing U^\dagger effectively transforms $|i\rangle$ into the desired state:

$$U^\dagger |i\rangle = \sum_j |\phi_j\rangle \langle j|i\rangle = |\phi_i\rangle. \quad (3.52)$$

This strategy for conducting measurements in non-standard bases finds a notable application in the Bell measurement, a topic we shall delve into in § 8.4.

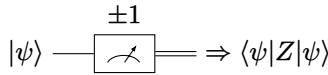
3.4.6 * Sampling Errors

In quantum computing, we often execute a quantum circuit repeatedly to obtain the empirical expectation values of certain observables (see § 1.5.3). Due to the finite number of runs for the circuit, empirical values may deviate from theoretical expectation values. This deviation is known as the sampling error.

In this subsection, we will use the Pauli operators (X , Y , Z) to demonstrate this important aspect of quantum measurement. The Pauli operators are unique in that they are both unitary and Hermitian, enabling them to serve as both quantum gates and observables.

1 Z Measurement

A measurement in the computational basis is equivalent to measuring the Z observable. In this scenario, the readout can be assigned a value of $+1$ for $|0\rangle$ and -1 for $|1\rangle$, as illustrated in the subsequent diagram. Through a collection of such readouts, we can calculate the empirical expectation value and variance of Z , and estimate the sampling error, as will be discussed next.



2 Expected Value

Assume a quantum state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. The theoretical expectation value for the Z observable is

$$\langle Z \rangle \equiv \langle \psi | Z | \psi \rangle = |\alpha|^2 - |\beta|^2 = 2|\alpha|^2 - 1, \quad (3.53)$$

since $|\alpha|^2 + |\beta|^2 = 1$.

To compute this empirically, the circuit is run N times with each measurement readout $z_i \in \{\pm 1\}$. The empirical expectation value is then:

$$\langle Z \rangle \approx \frac{1}{N} \sum_{i=1}^N z_i. \quad (3.54)$$

3 Standard Deviation

Sampling error pertains to the deviation of the empirical $\langle Z \rangle$ from the theoretical value. The standard deviation serves as an estimate for this error.

The theoretical variance of the observable Z in the state $|\psi\rangle$ is given by

$$(\Delta Z)^2 = \langle Z^2 \rangle - \langle Z \rangle^2. \quad (3.55)$$

Given that $\langle Z^2 \rangle = |\alpha|^2 + |\beta|^2 = 1$, the theoretical variance for the Z observable in this state becomes

$$(\Delta Z)^2 = 1 - \langle Z \rangle^2 = 1 - (|\alpha|^2 - |\beta|^2)^2 = 4|\alpha\beta|^2. \quad (3.56)$$

And consequently, the theoretical standard deviation is:

$$\Delta Z = 2|\alpha\beta|. \quad (3.57)$$

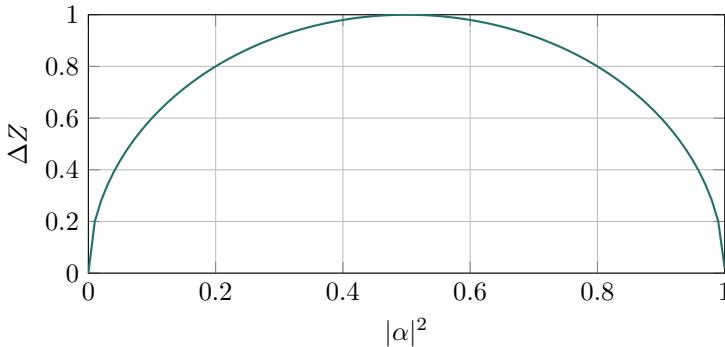


Figure 3.5: Plot of ΔZ as a function of $|\alpha|^2$

The relationship between ΔZ and $|\alpha|^2$ is depicted in Fig. 3.5. This curve elucidates the statistical nature of quantum measurement. When $|\alpha|^2 = 0.5$, the standard deviation ΔZ reaches its maximum value of 1, while the mean $\langle Z \rangle$ is zero. In addition, when $|\psi\rangle$ is a basis state—corresponding to cases where either $\alpha = 0$ or $\beta = 0$ —the standard deviation ΔZ is zero. This behavior is in agreement with the established principle that basis states can be measured with zero uncertainty.

4 Sampling Error of the Mean

The sampling error for the expected value $\langle Z \rangle$ is usually characterized by the standard error of the mean (SEM), expressed as:

$$\text{SEM}_{\langle Z \rangle} = \frac{\Delta Z}{\sqrt{N}}, \quad (3.58)$$

where ΔZ is the standard deviation of the individual measurements and N is the total number of measurements. Thus,

$$\text{SEM}_{\langle Z \rangle} = \frac{2|\alpha\beta|}{\sqrt{N}}. \quad (3.59)$$

As the number of measurements N becomes large, the standard error of the mean (SEM) approaches zero. This indicates that the empirical average $\langle Z \rangle$ converges to the true expected value.

Exercise 3.21 Consider a collection of N qubits, each in the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. If the goal is to measure the expected value $\langle Z \rangle$ with a 99% accuracy, what is the minimum value of N required to achieve this level of precision?

5 X Measurement

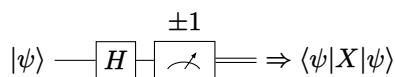
Measuring the X observable corresponds to making measurements in the $\{|+\rangle, |-\rangle\}$ basis. If the quantum state is expressed as $|\psi\rangle = a|+\rangle + b|-\rangle$, the equations developed for Z measurement remain applicable. However, the coefficients a and b are related to α and β via:

$$a = \frac{1}{\sqrt{2}}(\alpha + \beta), \quad b = \frac{1}{\sqrt{2}}(\alpha - \beta). \quad (3.60)$$

The relationship between a and b with α and β allows us to translate the expected values and standard deviations calculated in the $\{|+\rangle, |-\rangle\}$ basis to the computational basis, and vice versa.

Exercise 3.22 Derive formulas for the theoretical standard deviation (ΔX) for the X observable in terms of α and β .

As elaborated in § 3.4.4, measuring in the $\{|+\rangle, |-\rangle\}$ basis can be done by transforming the quantum state with the basis-change unitary operator (here the Hadamard operator H) before measuring in the computational basis. Through repeated readouts, we can calculate the empirical expectation value X and estimate its sampling error, similar to our preceding discussion for Z .



3.4.7 * Advanced Topics in Quantum Measurements

The preceding quantum measurement framework can be extended to include more complex scenarios such as POVMs, weak measurements, and nondestructive measurements, which allow for greater flexibility and applicability in quantum information science.

POVM: In quantum information and computation, POVMs are often used to model more complex measurement scenarios, including noisy or partial measurements. They are particularly useful in the context of open quantum systems, where they describe the various possible outcomes of measurements on systems interacting with their environment.

Weak Measurement: Traditional measurement collapses the state vector into one of the eigenstates of the observable. Weak measurements, on the other hand, only slightly perturb the quantum state, yielding only partial information about the observable. Weak measurements can be understood as a statistical limit of a series of less invasive measurements.

Nondestructive Measurements: The Measurement Postulate implies that a measurement fundamentally changes the state of the system. Nondestructive measurements try to sidestep this by obtaining information about the quantum state without significantly altering it, which is particularly useful in quantum error correction and certain quantum computation protocols.

The following texts are recommended for readers interested in these topics:

- Kurt Jacobs. *Quantum Measurement Theory and its Applications*. English. 1st. Cambridge University Press, 2014, page 554. ISBN: 978-1-107-02548-6.
- Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2017. ISBN: 978-1-107-17616-4.

3.5 *Application to Quantum State Tomography

The topic of this section is quantum state tomography, a powerful technique used to reconstruct the state of a quantum system. We apply the frameworks developed in this chapter on quantum states, basis transformation, and quantum measurements to this process, providing a cohesive case study.

Quantum state tomography involves reconstructing unknown quantum states based on measurement outcomes. As an illustrative example, let's assume we have $N = 2000$ qubits, all in an identical, yet unknown state, $|\psi\rangle$. These qubits are divided equally between Alice and Bob, with each receiving $N_a = 1000$ qubits. Alice measures her qubits in the computational basis, $\{|0\rangle, |1\rangle\}$, and observes $N_{a0} = 241$ qubits in the $|0\rangle$ state. Conversely, Bob uses the $\{|+\rangle, |-\rangle\}$ basis and detects $N_{b+} = 852$ qubits in the $|+\rangle$ state.

3.5.1 *Extracting Parameters from Alice's Measurements

From § 3.2, we know that a general qubit state expressed in the $\{|0\rangle, |1\rangle\}$ basis can be represented as:

$$|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + \sin \frac{\theta}{2} e^{i\phi} |1\rangle. \quad (3.61)$$

Given this, the measurement probabilities for the states $|0\rangle$ and $|1\rangle$ are, respectively, $P_0 = \cos^2 \frac{\theta}{2}$ and $P_1 = \sin^2 \frac{\theta}{2}$. Using these probabilities, θ can be deduced from the difference $P_0 - P_1$:

$$\cos \theta = P_0 - P_1. \quad (3.62)$$

From Alice's data, we deduce:

$$P_0 \approx \frac{241}{1000}, \quad P_1 \approx \frac{1000 - 241}{1000}. \quad (3.63)$$

This results in an approximate value of $\theta \approx 2.12$ radians.

3.5.2 *Extracting Parameters from Bob's Measurements

Given the state in the $\{|+\rangle, |-\rangle\}$ basis, the expression becomes:

$$|\psi\rangle = \frac{\cos \frac{\theta}{2} + e^{i\phi} \sin \frac{\theta}{2}}{\sqrt{2}} |+\rangle + \frac{\cos \frac{\theta}{2} - e^{i\phi} \sin \frac{\theta}{2}}{\sqrt{2}} |-\rangle. \quad (3.64)$$

From this expression, using the difference in probabilities $P_+ - P_-$, we can derive $\cos \phi \sin \theta$:

$$\sin \theta \cos \phi = P_+ - P_-. \quad (3.65)$$

Exercise 3.23 Derive Eq. 3.65 starting from Eq. 3.64.

From Bob's observations, we gather:

$$P_+ \approx \frac{852}{1000}, \quad P_- \approx \frac{1000 - 852}{1000}. \quad (3.66)$$

With θ previously derived from Alice's measurements, we can calculate ϕ using Eq. 3.65, arriving at $\phi \approx 0.604$.

3.5.3 *Reconstructing the Quantum State

With the deduced values of θ and ϕ , the original quantum state can be reconstructed using Eq. 3.61:

$$|\psi\rangle \approx 0.491 |0\rangle + 0.871 e^{0.604i} |1\rangle. \quad (3.67)$$

Exercise 3.24 *Sampling Errors (See § 3.4.6)

Quantum measurements inherently exhibit stochastic behavior, leading to only approximate acquisition of P_0 , P_1 , P_+ , and P_- through repeated measurements. This phenomenon is termed sampling error. The precision of these values is related to the sample size N , i.e., the number of measurements. Such sampling intricacies are endemic to quantum computing.

- (a) Deduce the average sampling errors for P_0 , P_1 , P_+ , and P_- in terms of N , N_a , N_b , etc.
- (b) Calculate the average sampling errors for θ and ϕ .
- (c) For the illustrated scenario, how large should N be to ensure the errors in θ and ϕ remain within a 0.1% threshold?

3.6 Summary and Conclusions

Comprehensive Framework for Qubits and Qudits

In the current chapter, we embarked on a journey to build a comprehensive framework for qubits and qudits. Serving as the backbone of quantum computing and quantum information systems, qubits and their generalized counterparts, qudits, are integral to our understanding of quantum phenomena. The chapter's aim was to transcend the limitations of specific examples like photons and electron spins, and to present a universal model that caters to the broad spectrum of qubit manifestations.

The topics of change of basis and quantum measurement theory were given special emphasis, not only for their theoretical significance but also for their practical applications in quantum computing. This chapter's examination of quantum state tomography illustrated the tangible applications of the aforementioned principles, highlighting their relevance in real-world scenarios.

Physical Manifestations and Mathematical Models

The diverse platforms that are actively being researched for qubit implementation underscore the versatility and adaptability inherent to quantum systems. We presented an overview of these platforms, drawing attention to their distinct methods of information encoding and their specific basis states.

Despite these various physical representations, a unified mathematical model for qubits emerged, anchored firmly in the tenets of quantum mechanics. This mathematical abstraction enables us to draw parallels across different qubit systems, underscoring the universal principles that underpin them.

Basis Rotations and Unitary Transformations

A considerable portion of this chapter was allocated to understanding quantum states and operators across diverse bases. The importance of bases in quantum mechanics was likened to reference frames in classical physics, offering readers a familiar touchstone. At the heart of our discourse is the concept that changing a basis aligns with a unitary transformation. We explored the transformation of vectors, operators, inner products, and operator averages in-depth.

This detailed inspection of basis changes proves invaluable in two key areas: firstly, alternate bases often shed light on nuanced insights into quantum states, proving essential for understanding quantum operations or algorithms. Secondly, operations triggered by quantum gates or other unitary transformations can be construed as a change of basis within the associated vector spaces. While both scenarios revolve around a change of basis, the nuances distinguishing them are subtle.

A General Formulation for Quantum Measurement

Our dialogue transitioned towards a generalized formulation of quantum measurements, rooted in the observable and measurement postulates of quantum mechanics. This discussion sets the stage for future explorations into decoherence, as well as advanced measurement strategies like POVM and weak measurements prevalent in contemporary quantum computing and quantum information contexts.

A salient insight was recognizing that the observable-based measurements, detailed in previous chapters, represent a specific case within the expansive framework of quantum measurement: the projective measurement.

Additionally, we embarked on a comprehensive examination of measurements across varying bases, augmenting our understanding of the quantum observation process through diverse reference frames.

Upcoming Topics

Having established a robust foundation on qubits and qudits, their physical embodiments, and the mathematical structures characterizing them, we are poised to explore the dynamics of quantum systems. The forthcoming chapter will address the

temporal evolution of quantum states. Governed by the Schrödinger equation, the dynamics of qubit systems holds the key to understanding the mechanisms by which quantum information is processed, manipulated, and observed, especially through the lens of quantum gates and circuits.

Our journey will initially concentrate on the dynamics of a single qubit. This will set the stage to navigate the complexities of multi-qubit systems, entanglements, and interactions. These foundational discussions will be instrumental as we transition to more advanced realms, such as quantum algorithms, quantum error correction, and the architectural design of extensive quantum systems.

Problem Set 3

- 3.1** Consider a unitary operator U with an eigenvector $|\psi\rangle$ such that $U|\psi\rangle = e^{2\pi i \theta} |\psi\rangle$.

The phase estimation algorithm applies U^{2^j} on $|\psi\rangle$ for different values of $j = 0, 1, 2, \dots$. Your task here is to compute $U^{2^j} |\psi\rangle$.

- 3.2** (Computational basis to general basis transformation.) Derive the matrix U for basis change $\{|0\rangle, |1\rangle\} \rightarrow \{|\psi\rangle, |\psi_{\perp}\rangle\}$ where $|\psi\rangle$ is defined by Eq. 3.2 and $|\psi_{\perp}\rangle$ by Eq. 3.5. Then, using U and using inner product expressions, determine the representation of $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$ in the new basis.

Hint: $U = |\psi\rangle\langle 0| + |\psi_{\perp}\rangle\langle 1|$.

- 3.3** Consider a four-level quantum system (qudit with $d = 4$). The following orthonormal basis states span the state space \mathbb{C}^4 :

$$|\phi_0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |3\rangle), \quad (3.68a)$$

$$|\phi_1\rangle = \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle), \quad (3.68b)$$

$$|\phi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |3\rangle), \quad (3.68c)$$

$$|\phi_3\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle). \quad (3.68d)$$

- (a) Verify these basis states are orthonormal given that $\{|0\rangle, \dots, |3\rangle\}$ are.
- (b) Determine the transformation operator between the computational basis and the new basis in both bra-ket and matrix notation.

- 3.4** The transformation from the computational basis to the four-dimensional Hadamard basis is given by

$$H_4 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \quad (3.69)$$

- (a) Find the basis states of the Hadamard basis $\{|h_i\rangle\}$.
- (b) Calculate the representation of $|\psi\rangle = \sum_i c_i |i\rangle$ in the Hadamard basis, where $c_i \in \mathbb{C}$.
- (c) Calculate the representation of $|\psi\rangle = \sum_i c_i |h_i\rangle$ in the computational basis, where $c_i \in \mathbb{C}$.

3.5 Design a setup for measuring the Y observable, analogous to the setup provided for the X measurement detailed in § 3.4.6.5.

Hint: Consider the unitary transformation needed to map the eigenstates of Y to the computational basis.

3.6 Consider a collection of independent qubits all in the identical state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$.

- (a) Suppose $|\alpha|^2 = 0.99$. What is the probability of measuring 10 qubits and find all of them in $|0\rangle$? What is the probability of measuring 10 qubits and find 9 of them in $|0\rangle$ and 1 in $|1\rangle$?
- (b) Suppose we measure 10 qubits and find each in the state $|0\rangle$. Estimate the probability for $|\alpha|^2 > 0.99$, assuming $|\alpha|^2$ has a uniform distribution.

3.7 Derive formulas for the theoretical expectation value ($\langle Y \rangle$) and standard deviation (ΔY) for the Pauli- Y observable in terms of α and β for the qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$.

3.8 Two orthonormal bases in \mathbb{C}^d , denoted by $\{u_j\}$ and $\{v_j\}$, are termed mutually unbiased if all vector pairs, one from each basis, possess inner products with magnitude $\frac{1}{\sqrt{d}}$. That is, $|\langle u_j | v_k \rangle| = \frac{1}{\sqrt{d}}$. Determine a basis mutually unbiased to the computational basis for $d = 2$ and $d = 4$.

4. Dynamics of Quantum Systems

Contents

4.1	The Evolution Postulate of Quantum Mechanics	83
4.1.1	Properties of Unitary Transformation	84
4.2	* The Schrödinger Equation	84
4.2.1	* The Time-independent Schrödinger Equation	85
4.2.2	* The Time-dependent Schrödinger Equation	86
4.2.3	* Trotterization	87
4.3	Stationary Nature of Energy Eigenstates	88
4.3.1	Time Evolution Operator in the Basis of Energy Eigenstates	88
4.3.2	Stationary Energy Eigenstates	88
4.3.3	The Ground State and Excited States	89
4.3.4	Qubit as a Two-Level System	90
4.4	Universal Quantum Computing and Annealing	91
4.4.1	Gate-based Universal Quantum Computing	91
4.4.2	Adiabatic Quantum Computation and Quantum Annealing	92
4.5	* Larmor Precession and Rabi Oscillations	93
4.5.1	* Larmor Precession	93
4.5.2	* Rabi Oscillations	96
4.6	* Further Exploration	99
4.6.1	* Time Evolution Equations for Relativistic Systems	100
4.6.2	* Open Quantum Systems	100
4.7	* Deferred Proofs	101
4.8	Summary and Conclusions	103
	Problem Set 4	104

In the preceding chapters, we have explored how a qubit's state is represented by its state vector. As we extend our discussion to include the dynamics of qubits, we encounter the question of how a qubit's state evolves over time. Such evolution can result from various interactions, like the application of a quantum gate or quantum annealing processes. These interactions drive the changes we observe over time, which are described as the dynamics of quantum systems.

In the realm of quantum mechanics, dynamics pertains to the evolution of quantum states as a result of physical interactions. The equations and principles of quantum mechanics provide the framework for understanding this evolution. They allow us to predict and analyze how quantum states change, interact, and develop over time, which is fundamental to both the theoretical understanding and practical applications of quantum physics.

The dynamics of a quantum system is principally governed by the Evolution Postulate of quantum mechanics, a fundamental theorem dictating how quantum states evolve temporally. For qubit systems, this evolution is predominantly governed by the Schrödinger equation. Mastery of this fundamental equation is paramount to comprehending the intricate mechanisms by which quantum information is processed, manipulated, and observed in a wide array of quantum systems.

In this chapter, the exploration will introduce the general concepts of the dynamics of quantum system, using examples of a single qubit. Subsequent chapters will broaden this examination to encompass systems of multiple qubits, probing their interactions, entanglements, and the means by which their dynamics can be both controlled and exploited for potent computational tasks. This foundational comprehension of the dynamics of qubits will act as a nexus to more advanced subjects such as quantum algorithms, quantum error correction, and the architecture of large-scale quantum systems.

4.1 The Evolution Postulate of Quantum Mechanics

The Time Evolution Postulate is a fundamental principle in quantum mechanics, describing how quantum states change over time.

Postulate 4: Time Evolution

The time evolution of the state of a closed quantum system is governed by a unitary transformation.

Mathematically, we can express this Postulate as:

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle, \quad (4.1)$$

where $|\psi(t)\rangle$ is the quantum state of the system at time t , $U(t)$ is the time-evolution operator that describes how the state at time $t = 0$ evolves to the state at time t , and $|\psi(0)\rangle$ is the initial state of the system.

The necessity of the time-evolution operator being unitary in a closed quantum system stems from the conservation of total probability, which must remain equal to one. Essentially, a quantum state vector must always be normalized since its components represent probability amplitudes, and their square sum must equal one. As the quantum state naturally evolves over time, this normalization must be preserved. Unitary transformations fulfill this requirement by maintaining the inner product, thus ensuring the coherence and consistency of the behavior of the system.

4.1.1 Properties of Unitary Transformation

Unitary transformations are fundamental to quantum mechanics as they govern the evolution of the states of closed quantum systems. They are also essential in quantum computing, possessing several key properties that enable quantum state manipulation. For simplicity, we will denote $U(t)$ as U in the following discussion.

1. Preservation of Inner Products: The overlap of two quantum states, $|\phi\rangle$ and $|\psi\rangle$, remains invariant under the same unitary evolution. Explicitly,

$$\langle U\phi|U\psi\rangle = \langle\phi|U^\dagger U|\psi\rangle = \langle\phi|\psi\rangle. \quad (4.2)$$

A consequence of this property is that an orthonormal basis remains orthonormal under a unitary evolution.

Moreover, as $|\langle\phi|\psi\rangle|^2$ represents the probability of measuring state $|\phi\rangle$ in state $|\psi\rangle$, or vice versa, such probabilities remain unchanged under unitary evolution, preserving the probabilistic interpretation of quantum mechanics.

2. Reversibility: Unitary transformations are reversible, meaning that they can be undone by applying the inverse transformation using U^{-1} .
3. Linearity: Superposition principle holds during the unitary evolution. A unitary transformation applied to a superposition of quantum states results in a superposition of transformed states.
4. Unit Eigenvalues: The eigenvalues of a unitary transformation are complex numbers with an absolute value of 1.
5. Orthonormal Row and Column Vectors: The column vectors of a unitary matrix form an orthonormal set, and the same applies to the row vectors when transposed.

Unitary transformations can be *interpreted* as a change of basis in Hilbert space, effectively rotating the quantum state vector while preserving the structure of the space. In particular, quantum gates apply unitary transformations to quantum states. While computations are commonly performed in the computational basis, quantum states exist in a Hilbert space representable in any orthonormal basis. See § 3.3.2 for an in-depth discussion on the change of basis concept.

4.2 *The Schrödinger Equation

The Schrödinger equation is a cornerstone of non-relativistic quantum mechanics, governing the time evolution of quantum states in systems like electrons in atoms, quantum dots, and superconducting qubits. This equation links the time-dependent evolution of a quantum state to the system's Hamiltonian operator, often simply referred to as the Hamiltonian, H . The Hamiltonian represents the total energy observable of the system and thus acts as a crucial connection between the abstract quantum state and measurable physical quantities, like energy levels.

The Schrödinger equation is expressed as:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H(t) |\psi(t)\rangle, \quad (4.3)$$

where $|\psi(t)\rangle$ is the quantum state vector at time t , and \hbar is the reduced Planck constant.

While we will not delve into solving this equation here, we will focus on key concepts and results essential to understanding quantum gates, quantum hardware, and providing a foundation for further study.

Hermitian Property of the Hamiltonian

As a consequence that the solution of the Schrödinger equation must represent a unitary evolution given by Eq. 4.1, the Hamiltonian H must be a Hermitian operator. We will provide a proof of this property at the end of this section (see § 4.7).

From a physical point of view, the Hermitian condition for the Hamiltonian is necessarily true, since H represents the total system energy, and as such, its eigenvalues must be real numbers.

4.2.1 *The Time-independent Schrödinger Equation

In the simplest case where the Hamiltonian H does not change with time, the Schrödinger equation is expressed as:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle. \quad (4.4)$$

This equation has a general solution of the form:

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar} H t} |\psi(0)\rangle, \quad (4.5)$$

which can be verified by differentiating both sides over t (see § 4.7).

The exponential term

$$U(t) = e^{-\frac{i}{\hbar} H t} \quad (4.6)$$

is the unitary time-evolution operator, and it determines how the quantum state $|\psi(0)\rangle$ at the initial time evolves to the state $|\psi(t)\rangle$ at time t .

Here the exponential function of operator A is defined as

$$e^A \equiv \exp(A) \equiv \sum_{n=0}^{\infty} \frac{1}{n!} A^n, \quad (4.7)$$

where A is a normal operator, meaning $AA^\dagger = A^\dagger A$. Normal operators include Hermitian and unitary operators.

Properties of the Exponential Operator

The exponential operator e^A has several important properties:

Product Property

In general, $e^{A+B} \neq e^A e^B$. However,

$$e^{A+B} = e^A e^B, \quad \text{if } [A, B] = 0. \quad (4.8)$$

We will provide a proof of this property at the end of this section (see § 4.7).

Unitary Property

If A is Hermitian, then e^{iA} is unitary. A proof is outlined below:

$$(e^{iA}) (e^{iA})^\dagger = e^{iA} e^{-iA^\dagger} \quad (4.9a)$$

$$= e^{iA} e^{-iA} \quad (4.9b)$$

$$= e^{iA-iA} = e^0 = I. \quad (4.9c)$$

This property confirms that $U(t)$ in Eq. 4.6 is unitary, consistent with the Time Evolution Postulate that the state of a closed quantum system is governed by a unitary transformation.

Generalized Euler Formula

Consider the case where $A^2 = I$ (i.e., A is involutory) and θ is a real number. In this scenario, the following relationship holds:

$$e^{i\theta A} = \cos \theta I + i \sin \theta A. \quad (4.10)$$

This equation can be analogized to the Euler formula, $e^{i\theta} = \cos \theta + i \sin \theta$. A detailed proof of this property is provided at the end of this section (see § 4.7). Readers are encouraged to study the mathematics behind this proof, as the techniques and concepts involved are highly relevant to the field of quantum computing.

Exercise 4.1 Express e^{iX} as a single 2×2 matrix, where X is the Pauli- X matrix.

Exercise 4.2 * Express $e^{i(I+Z)/2}$ to a single 2×2 matrix, where Z is the Pauli- Z matrix and I is the identity matrix.

4.2.2 *The Time-dependent Schrödinger Equation

When the Hamiltonian varies with time, denoted as $H(t)$, the formulation of the time-evolution operator becomes more complex. For cases where $H(t_1)$ commutes with $H(t_2)$ for any t_1 and t_2 within the time domain of interest, the solution to the time-dependent Schrödinger equation is given by

$$U(t) = \exp \left(-\frac{i}{\hbar} \int_0^t H(t') dt' \right). \quad (4.11)$$

In scenarios where $H(t_1)$ and $H(t_2)$ do not commute, the solution is expressed using a time-ordered exponential, known as the Dyson series:

$$U(t) = \mathcal{T} \exp \left(-\frac{i}{\hbar} \int_0^t H(t') dt' \right). \quad (4.12)$$

Here, \mathcal{T} represents the time-ordering operator, which organizes the terms in the exponential based on their time arguments, arranging factors corresponding to later times to the left of those associated with earlier times.

It can be shown ([8]) that the time-evolution operator $U(t)$ is unitary even when $H(t)$ is time-dependent, as it preserves the inner product and the norm of

state vectors. This ensures that the probabilities remain well-defined and conserved throughout the evolution.

While the Dyson series is fundamental to a comprehensive understanding of time-dependent quantum mechanics, it is not crucial for a general quantum computing audience. For practical applications in quantum computing and simulation, the Trotterization approximation is often more relevant.

4.2.3 *Trotterization

Trotterization, also known as the Trotter-Suzuki decomposition, is a method used in quantum mechanics to approximate the evolution of quantum systems with non-commuting terms in their Hamiltonians. This technique is particularly useful in quantum computing for simulating complex quantum systems, where exact solutions are often computationally infeasible.

When dealing with a time-varying Hamiltonian $H(t)$, the Trotterization formula can be used to approximate the time-evolution operator $U(t)$. Consider dividing the total evolution time T into N small intervals, $\Delta t = T/N$. The Trotterization formula for a time-dependent Hamiltonian is:

$$U(T) \approx \prod_{n=1}^N \exp\left(-\frac{i}{\hbar}H(t_n)\Delta t\right), \quad (4.13)$$

where $t_n = n\Delta t$ represents the time at the n th interval. This product is ordered such that factors corresponding to later times are placed to the left of those associated with earlier times.

This formula approximates the time-ordered exponential (Eq. 4.12) by assuming the Hamiltonian remains approximately constant over each interval Δt . Each term in the product is an exponential of the Hamiltonian evaluated at a specific time t_n , scaled by Δt . The accuracy of this approximation increases as $N \rightarrow \infty$ (i.e., as $\Delta t \rightarrow 0$).

Trotterization is also useful for Hamiltonians containing non-commuting terms, common in many real-world quantum systems. For example, consider a Hamiltonian of the form $H = H_1 + H_2$, even when time-independent. In such cases, the exact solution to the Schrödinger equation can be challenging to compute, as $e^{i(H_1+H_2)} \neq e^{iH_1}e^{iH_2}$. Trotterization allows us to approximate the time-evolution operator over a small interval Δt :

$$U(T) \approx \left(\exp\left(-\frac{i}{\hbar}H_1\Delta t\right) \exp\left(-\frac{i}{\hbar}H_2\Delta t\right) \right)^N. \quad (4.14)$$

The effectiveness of Trotterization depends on the size of Δt and the nature of the Hamiltonian. For rapidly changing or highly non-commutative Hamiltonians, more sophisticated versions, such as higher-order Suzuki-Trotter expansions, might be necessary.

Exercise 4.3 Consider a Hamiltonian $H = \alpha X + \beta Z$, where X and Z are the Pauli matrices, and $\alpha, \beta \in \mathbb{R}$. Write an approximate solution to the Schrödinger equation using Trotterization.

4.3 Stationary Nature of Energy Eigenstates

As a consequence of the time-independent Schrödinger equation, energy eigenstates are stationary, as explained below. This property plays an important role in many areas of quantum physics and quantum computing.

Let $|E_k\rangle$ be the eigenvectors of the Hamiltonian operator H , with the corresponding energy eigenvalues E_k . Since H is Hermitian, E_k has real values. When these eigenvalues are discrete, they are commonly referred to as the energy levels of the system.

4.3.1 Time Evolution Operator in the Basis of Energy Eigenstates

According to the spectral decomposition theorem in linear algebra, H can be expressed as:

$$H = \sum_k E_k |E_k\rangle\langle E_k|. \quad (4.15)$$

Each term in the sum projects the state vector onto the k -th energy eigenstate $|E_k\rangle$ and scales it by the corresponding energy eigenvalue E_k .

The time-evolution operator $U(t)$ can be expressed as:

$$U(t) = e^{-\frac{i}{\hbar}Ht} = \sum_k e^{-\frac{iE_k t}{\hbar}} |E_k\rangle\langle E_k|. \quad (4.16)$$

This equation describes the time-evolution operator $U(t)$ as a sum over all energy eigenstates with the corresponding energy eigenvalues. Each term in the sum evolves the k -th energy eigenstate $|E_k\rangle$ by a phase factor $e^{-\frac{iE_k t}{\hbar}}$. This equation can be viewed as the solution to the time-independent Schrödinger equation in the basis of the energy eigenstates.

4.3.2 Stationary Energy Eigenstates

Equation 4.16 shows that if a system is initially in an energy eigenstate $|\psi(0)\rangle = |E_k\rangle$, then its time evolution is simply a phase factor multiplying the initial state:

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle = e^{-\frac{iE_k t}{\hbar}} |E_k\rangle. \quad (4.17)$$

This means that if a system is prepared in an energy eigenstate, it will remain in that state with a time-dependent phase factor. Such a state is called a stationary state. Stationary states are quantum states that do not change with time, except for an overall phase factor.

If the system is not in an energy eigenstate initially, then its time evolution will involve a superposition of energy eigenstates, and the amplitudes of these energy eigenstates will evolve in time according to the corresponding phase factors. For example, say the system is initially in a superposition of $|E_0\rangle$ and $|E_1\rangle$:

$$|\psi(0)\rangle = \alpha |E_0\rangle + \beta |E_1\rangle. \quad (4.18)$$

At time t , the system state $|\psi(t)\rangle$ is given by

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle = e^{-\frac{iE_0t}{\hbar}}\alpha|E_0\rangle + e^{-\frac{iE_1t}{\hbar}}\beta|E_1\rangle. \quad (4.19)$$

It's evident that the probabilities of finding the system in either energy state remain constant, as $|e^{-\frac{iE_0t}{\hbar}}\alpha|^2 = |\alpha|^2$ and $|e^{-\frac{iE_1t}{\hbar}}\beta|^2 = |\beta|^2$.

The above discussion highlights the stationary nature of energy eigenstates in an ideal quantum system, where the probability distributions of these states remain constant over time, assuming the system is perfectly isolated and not subjected to any external forces or interactions.

However, in practice, real quantum systems are never completely closed or isolated. They invariably interact with their surroundings to some extent, leading to deviations from the idealized stationary behavior. These interactions can cause transitions between different energy states and introduce phenomena such as decoherence and dissipation, significantly influencing the system's behavior.

Exercise 4.4 Consider a Hamiltonian H :

$$H = X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

- (a) Diagonalize this Hamiltonian to find its eigenvalues and eigenvectors.
- (b) Given the time-evolution operator $U(t) = e^{-iHt}$, compute $U(t)$ for a time t , using the eigendecomposition of H .
- (c) If we start with a state $|0\rangle$, and we evolve it using $U(t)$, the resulting state will be $U(t)|0\rangle$. Compute the probabilities of measuring $|0\rangle$ and $|1\rangle$ after this evolution.

4.3.3 The Ground State and Excited States

The ground state, $|E_0\rangle$, represents the lowest energy state the system can occupy. In non-ideal, real-world quantum systems, this state is typically the most stable configuration, often realized as the system's natural state in the absence of external energy inputs. It serves as the foundational reference point for defining other possible states of the system, known as the excited states. These excited states, having higher energy levels than the ground state, can be populated due to environmental interactions or external stimuli, highlighting the dynamic nature of quantum systems in realistic settings.

An excited state, $|E_1\rangle$ for example, is any state of a system that has a higher energy than the ground state. A qubit or particle has to absorb energy of the amount $E_1 - E_0$ to be “excited” from the ground state to this less stable state. This energy can be supplied through various means, such as by applying an external electric or magnetic field, or by exposing the system to electromagnetic radiation (e.g., microwave or laser) of angular frequency

$$\omega = \frac{E_1 - E_0}{\hbar}. \quad (4.20)$$

The excitation energy can also be provided, usually unintendedly, through thermal energy, which is why some quantum hardware operate at low temperatures.

A superposition state, which is a combination of several possible states, can be achieved through careful control of the energy input and the system's interactions.

An excited state or superposition state is inherently unstable because it represents a higher energy configuration of the system. When a system is in an excited state, it has multiple pathways to release the excess energy and return to the ground state. This process is known as relaxation. The mechanisms for relaxation can vary, including the emission of photons or phonons, collisions with other particles, or other interactions that allow the system to shed its excess energy. The tendency to return to the ground state also give rise to the short lifetime of qubits - referred to decoherence in quantum computing.

These concepts lay the foundation for understanding a wide range of quantum behaviors and are central to fields like quantum computing hardware, where control of ground and excited states enables the manipulation of qubits for computational purposes.

4.3.4 Qubit as a Two-Level System

A qubit is a two-level system, meaning it involves only two energy eigenstates: the ground state $|E_0\rangle$ and the first excited state $|E_1\rangle$. These states form the computational basis:

$$|0\rangle \equiv |E_0\rangle, \quad |1\rangle \equiv |E_1\rangle. \quad (4.21)$$

In quantum computing, qubits are typically initialized in the ground state $|0\rangle$. Quantum gates are designed to manipulate the qubits by inducing transitions between energy eigenstates or creating superpositions. The stationary nature of energy eigenstates ensures that, in an idealized and isolated quantum system, the qubits remain in their initial states until a quantum gate or external force acts upon them.

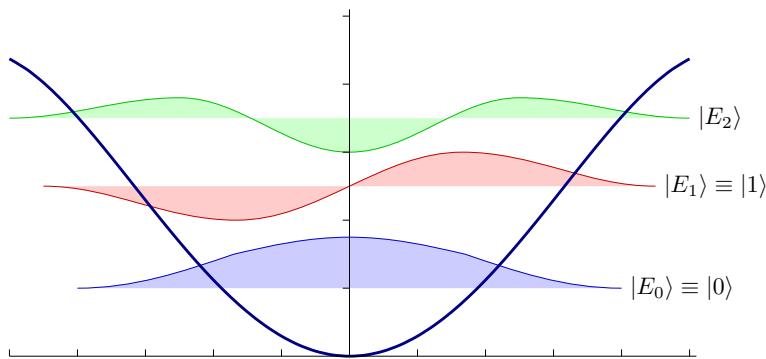


Figure 4.1: Example of a Practical Multi-Level System

In a practical quantum system, such as a superconducting qubit, energy levels are often more complex than a simple two-level system. These levels might include $|E_0\rangle$, $|E_1\rangle$, and $|E_2\rangle$, for example, where only $|E_0\rangle$ and $|E_1\rangle$ are actively used for the qubit operation, and the existence of $|E_2\rangle$ is considered undesirable.

To mitigate potential problems, the system is meticulously engineered so that the energy levels $|E_0\rangle$, $|E_1\rangle$, and $|E_2\rangle$ are unequally spaced, i.e., $E_1 - E_0 \neq E_2 - E_1$. This careful design ensures that energy specifically targeting the transition between $|E_0\rangle$ and $|E_1\rangle$ will not inadvertently affect the transition between $|E_1\rangle$ and $|E_2\rangle$. By minimizing the interaction with the undesirable level $|E_2\rangle$, the system gains greater stability and control, critical factors for the effective functioning of quantum computation. This is illustrated in Fig. 4.1.

Exercise 4.5 Given the Hamiltonian $H = \hbar\omega \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$, where $\omega \in \mathbb{R}$, find the energy levels E_0 and E_1 , and corresponding eigenstates $|E_0\rangle$ and $|E_1\rangle$.

4.4 Universal Quantum Computing and Annealing

In the rapidly advancing field of quantum computing, two predominant paradigms have emerged as foundational approaches to quantum computation: gate-based universal quantum computing and quantum annealing. While gate-based universal quantum computing is designed for a broad range of algorithms and can theoretically solve any computable problem, quantum annealing is particularly tailored for solving optimization problems by finding the minimum of a given objective function. The principle of unitary evolution of quantum states forms the foundation for these quantum computing paradigms, which we elucidate in the following subsections.

4.4.1 Gate-based Universal Quantum Computing

Gate-based quantum computers modulate qubit states using a series of predesigned unitary operations, known as quantum gates. Similar to classical digital computers where data processing occurs through logic gates, gate-based quantum computers process information by sequentially applying quantum gates to an initial state, iteratively evolving it into the final state that represents the computation result.

Quantum gates can be conceptualized as discrete counterparts to the continuous time-evolution operators in the Schrödinger equation. Implementing a quantum gate mandates a controlled alteration of the qubit's Hamiltonian such that the intended transformation applies to its state. This control could entail utilizing external elements like electromagnetic radiation (e.g., microwave pulses) or modulating other system-specific parameters (e.g., for superconducting qubits, trapped ions, or quantum dots).

Given a Hamiltonian H designed for the gate's function and an active time interval Δt , the unitary evolution operator $U(\Delta t)$, as informed by the Schrödinger equation, is represented as:

$$|\psi(\Delta t)\rangle = U(\Delta t) |\psi(0)\rangle. \quad (4.22)$$

Assuming the gate Hamiltonian H remains consistent over Δt , $U(\Delta t)$ can be articulated as:

$$U(\Delta t) = e^{-\frac{iH\Delta t}{\hbar}}. \quad (4.23)$$

Exercise 4.6 Consider a one-qubit quantum gate defined by the Hamiltonian $H = \frac{\hbar}{2}Z$. Derive the unitary evolution operator $U(\Delta t)$ over an interval Δt and elucidate the influence of this gate on an arbitrary qubit state.

4.4.2 Adiabatic Quantum Computation and Quantum Annealing

Adiabatic Quantum Computation (AQC) is based on the adiabatic theorem of quantum mechanics, wherein the gradually varied time evolution of a quantum system's Hamiltonian transitions the system from its initial to a desired final state.

1 Adiabatic Quantum Computation (AQC)

AQC leverages the inherent unitary evolution of quantum states governed by the time-dependent Schrödinger equation. According to the adiabatic principle, if a quantum system is initially in an eigenstate (usually the ground state) of a Hamiltonian and the Hamiltonian changes sufficiently slowly, and without encountering level crossings, the system will remain predominantly in the instantaneous eigenstate of the evolving Hamiltonian, avoiding transitions to other energy states. The probability of the system deviating from the desired eigenstate and ending up in other states decreases exponentially with the slowness of the change in the Hamiltonian.

The approach involves encoding a computational problem into a quantum system, where the ground state of this system corresponds to the problem's solution. The process starts with an initial, simple Hamiltonian, H_{initial} , with an easily identifiable ground state. This Hamiltonian then undergoes a slow (adiabatic) evolution to a final Hamiltonian, H_{target} , encoding the solution to the problem.

The Hamiltonian evolution is described by:

$$H(t) = (1 - s(t))H_{\text{initial}} + s(t)H_{\text{target}}, \quad (4.24)$$

where $s(t)$ is a scheduling parameter varying from 0 to 1 during the process, smoothly transitioning from H_{initial} to H_{target} .

Consider, for example, a spin qubit in a time-dependent magnetic field. The initial Hamiltonian $H_{\text{initial}} = Z$ has a ground state $|0\rangle$, corresponding to the spin aligned with the magnetic field in the z direction. As the field slowly rotates to the x direction, represented by $H_{\text{target}} = X$, the qubit's ground state evolves to $|+\rangle$, provided the change is adiabatic. In more complex systems with many interacting qubits, the final Hamiltonian H_{target} is not as straightforward and represents the complex problem to be solved.

2 Quantum Annealing: An Implementation of AQC

Quantum annealing is an implementation within the AQC framework. Unlike methods relying on quantum gates, it allows the system to naturally evolve under its quantum-mechanical dynamics. The term “annealing” refers to the process in which the Hamiltonian is adiabatically changed, akin to the metallurgical technique of slowly cooling materials to reach a stable state.

Implementing quantum annealing requires precise control over physical parameters characterizing the system's Hamiltonian. For instance, in superconducting quantum annealing systems, the Hamiltonian is implemented by adjusting magnetic

fields and couplings between qubits. By carefully tuning these parameters, the system is guided through the desired Hamiltonian evolution.

Quantum annealing computers are particularly effective for problems modeled by the Ising model, which can represent many optimization challenges. We will delve deeper into the Ising model in § 6.5.3 and explore its application in quantum annealing and other quantum algorithms in § 11.2.

4.5 *Larmor Precession and Rabi Oscillations

This section focuses on Larmor precession and Rabi oscillations, two cornerstone phenomena that shed light on the behavior of quantum systems subject to external fields. The dynamics of qubit systems is guided by the Schrödinger equation, which outlines their unitary evolution. Larmor precession and Rabi oscillations serve as archetypal cases to better grasp this evolution. By exploring these phenomena, we aim to demonstrate effective strategies for solving the Schrödinger equation. Their significance is further underscored by their relevance to quantum computing, especially in the practical aspects such as hardware design and quantum gate implementation.

4.5.1 *Larmor Precession

Larmor precession, or spin precession, occurs when a quantum particle with intrinsic angular momentum, or spin, is placed in an external magnetic field. The interaction between the particle's magnetic moment and the external field causes the particle's spin to precess, or rotate, around the direction of the magnetic field.

In the context of quantum computing, Larmor precession is often used as a mechanism to manipulate and control the state of a qubit. By applying the external magnetic field for a designed time interval, the qubit's spin can be precisely controlled, allowing for the implementation of specific quantum gates and operations. Larmor precession is also used in magnetic resonance imaging (MRI) to create images of the body's internal structures, and in nuclear magnetic resonance (NMR) spectroscopy to analyze the properties of molecules and materials.

1 Solution to the Schrödinger Equation

The Hamiltonian, representing the energy of interaction, of a spin with an external magnetic field B in the z -direction can be expressed as:

$$H = -\frac{\hbar\omega}{2}Z, \quad (4.25)$$

where ω is the Larmor frequency (see details explained at the end of this subsection).

The dynamics of the system is governed by the Schrödinger equation (Eq. 4.4), which has a solution in the form of the time-evolution operator $U(t)$ given by Eq. 4.6. Due to the property $Z^2 = I$ for Pauli matrix Z , we can simplify the matrix exponential in $U(t)$ into a regular matrix (refer to Eq. 4.10):

$$U(t) = e^{-iHt/\hbar} = e^{i\omega t Z/2} \quad (4.26a)$$

$$= \cos \frac{\omega t}{2} I + i \sin \frac{\omega t}{2} Z \quad (4.26b)$$

$$= \begin{bmatrix} e^{\frac{i\omega t}{2}} & 0 \\ 0 & e^{-\frac{i\omega t}{2}} \end{bmatrix}. \quad (4.26c)$$

If the spin is initially in the direction of (θ, ϕ) , its state vector is:

$$|\psi(0)\rangle = \begin{bmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\phi} \end{bmatrix}. \quad (4.27)$$

The time evolution of the system is given by:

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle \quad (4.28a)$$

$$= \begin{bmatrix} e^{\frac{i\omega t}{2}} & 0 \\ 0 & e^{-\frac{i\omega t}{2}} \end{bmatrix} \begin{bmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\phi} \end{bmatrix} \quad (4.28b)$$

$$= e^{\frac{i\omega t}{2}} \begin{bmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i(\phi - \omega t)} \end{bmatrix}, \quad (4.28c)$$

which is the analytic solution to the Schrödinger equation for Larmor precession.

2 Interpretation

From Eq. 4.28 we see that the azimuthal angle of the spin direction at time t is given by

$$\phi(t) = \phi - \omega t, \quad (4.29)$$

while the polar angle θ remains constant.

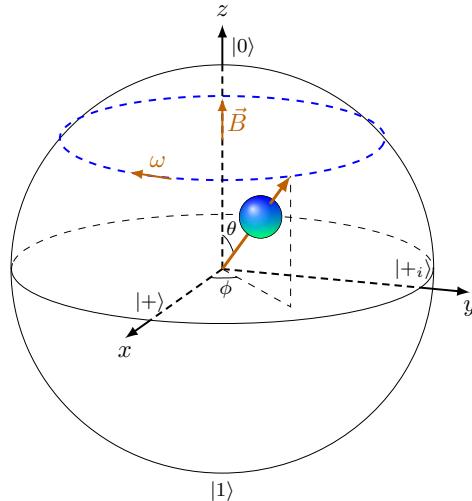


Figure 4.2: Larmor Precession on the Bloch Sphere

This can be visualized as the state vector of the spin rotating around the z -axis (the direction of the magnetic field), tracing out a circle on the surface of the Bloch

sphere, as illustrated in Fig. 4.2. The angular frequency of the rotation is the Larmor frequency ω . This type of motion is known as precession, hence the name Larmor precession.

Larmor precession can be likened to the behavior of a spinning top. Just as a top precesses around the vertical direction when not perfectly upright due the influence of gravity, a spin in a magnetic field precesses around the field direction if it's not aligned with the field. This precession is caused by the interaction between the spin's magnetic moment and the magnetic field, similar to the gravitational torque on the top. In the case of an open system, various interactions with the environment, often undesirable, can cause the spin to gradually align more closely with the magnetic field direction over time.

3 Implementation of the $R_z(\gamma)$ Gate

Larmor precession can be utilized to precisely control a spin-based qubit by aligning the magnetic field at an angle θ relative to the spin direction, and then activating the magnetic field for a time interval Δt . The spin direction will precess by an angle $\gamma = \omega\Delta t$. This is exactly the Pauli-Z rotation gate $R_z(\gamma)$, discussed in § 5.2.5.

Exercise 4.7 With the $|\psi(t)\rangle$ given in Eq. 4.28, calculate the expected values $\langle\psi(t)|X|\psi(t)\rangle$, $\langle\psi(t)|Y|\psi(t)\rangle$, and $\langle\psi(t)|Z|\psi(t)\rangle$.

4 *The Hamiltonian

The Hamiltonian for a spin in an external magnetic field \vec{B} , which is the interaction energy between the magnetic moment of the spin $\vec{\mu}$ and the magnetic field, is given by:

$$H = -\vec{\mu} \cdot \vec{B}. \quad (4.30)$$

The magnetic moment $\vec{\mu}$ of a spin is proportional to its spin angular momentum \vec{S} :

$$\vec{\mu} = \gamma \vec{S}, \quad (4.31)$$

In the case of a magnetic field aligned with the z -axis, the interaction energy becomes:

$$H = -\mu_z B = -\gamma B S_z. \quad (4.32)$$

The z -component of spin angular momentum, S_z , can be expressed in terms of the Pauli matrix Z as:

$$S_z = \frac{\hbar}{2} Z. \quad (4.33)$$

Combining these relations, we obtain:

$$H = -\frac{\hbar\gamma B}{2} Z = -\frac{\hbar\omega}{2} Z, \quad (4.34)$$

where $\omega = \gamma B$ is referred to as the Larmor frequency. This is the same as Eq. 4.25.

* A Clarification of the Sign of the Hamiltonian

In our discussion so far, we have focused on a generic exploration of Larmor precession involving a spin in a magnetic field oriented along the z -direction, assuming the spin particle is positively charged. For an electron, the Hamiltonian in Eq. 4.34 has an opposite sign: $H = \frac{\hbar\omega}{2} Z$. An astute reader might note a seeming inconsistency: the ground state for an electron in this scenario would be $|1\rangle$, not $|0\rangle$. To dispel any confusion, we delve into the nuances of different conventions in traditional quantum mechanics and quantum computing, as well as the distinctions between various physical systems.

- The lowest energy state, or ground state, corresponds to the magnetic moment being aligned with the magnetic field. For electrons, which possess a negative charge, this alignment translates to their spin being antiparallel to the magnetic field, traditionally referred to as spin-down in quantum mechanics. Conversely, for nuclei containing positively charged protons, the ground state corresponds to the spin being parallel to the magnetic field, or spin-up.
- In quantum computing, the state $|0\rangle$ is commonly associated with spin-up and is conventionally designated as the ground state. This choice is more a matter of computational convenience and is not necessarily consistent with all physical qubit systems.
- Furthermore, Larmor precession can be viewed as a perturbation to a base Hamiltonian that defines the ground state. This perspective is particularly relevant when Larmor precession serves as a model for the manipulation and control of qubit states. The application of a magnetic field (or other fields) introduces this perturbation, resulting in the characteristic precessional motion of the spin.

This clarification aims to bridge the gap between theoretical quantum mechanics and practical quantum computing conventions, enhancing the understanding of these fundamental concepts.

4.5.2 * Rabi Oscillations

Rabi oscillations are a phenomenon that occurs when a two-level quantum system (e.g., a qubit or spin-1/2 particle) is subjected to an oscillatory driving field. This phenomenon is named after the physicist Isidor Isaac Rabi, who first observed it in the context of atomic transitions. The oscillation occurs when the frequency of the external driving force is resonant with the energy difference between the two energy eigenstates. In this case, the probability of finding the electron in one of the energy levels oscillates periodically with time, completing one full oscillation within a Rabi cycle.

Rabi oscillations play a crucial role in many areas of quantum physics, such as atomic physics, quantum optics, and quantum computing. In the context of quantum computing, Rabi oscillations can be used to manipulate qubits by driving transitions between their computational basis states (e.g., $|0\rangle$ and $|1\rangle$). In this case, carefully controlled electromagnetic pulses can be applied to induce Rabi oscillations, which allow for precise control of the superposition state of the qubit.

1 Solution to the Schrödinger Equation

The Hamiltonian of the two-level quantum system (under the rotating wave approximation, see below) is given by

$$H = \frac{\hbar\Omega}{2} Y, \quad (4.35)$$

where Y is the Pauli- Y matrix and Ω is known as the Rabi frequency, which characterizes the strength of the coupling between the system and the external field.



Larmor precession and Rabi oscillations, while physically distinct phenomena, share mathematical similarities, particularly in their involvement of a single Pauli operator in the Hamiltonian.

In quantum computing, Larmor precession is often modeled to represent z -axis rotations of the qubit state, while Rabi oscillations are typically associated with x - and y -axis rotations. In our main discussion, we choose the Y Pauli operator in the Hamiltonian to demonstrate the creation of the $|+\rangle$ superposition state.

As a heads-up, it's noteworthy that the implementation of the most essential two-qubit gates, such as the CNOT and the ZZ gates, involves ZX and ZZ interactions. These interactions and their roles in multi-qubit systems will be explored in Chapters 6 and 7. The physical principles underlying these interactions are similar to those of Larmor precession and Rabi oscillations discussed here.

The dynamics of the system is governed by the Schrödinger equation (Eq. 4.4), which has a solution in the form of the time-evolution operator $U(t)$ given by Eq. 4.6. Due to the property $Y^2 = I$ of the Pauli matrix Y , we can simplify the matrix exponential in $U(t)$ into a regular matrix:

$$U(t) = e^{-iHt/\hbar} = e^{-i\Omega t Y/2} \quad (4.36a)$$

$$= \cos \frac{\Omega t}{2} I - i \sin \frac{\Omega t}{2} Y \quad (4.36b)$$

$$= \begin{bmatrix} \cos \frac{\Omega t}{2} & -\sin \frac{\Omega t}{2} \\ \sin \frac{\Omega t}{2} & \cos \frac{\Omega t}{2} \end{bmatrix}. \quad (4.36c)$$

In the case where the initial state is $|\psi(0)\rangle = |0\rangle$, we have:

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle \quad (4.37a)$$

$$= \begin{bmatrix} \cos \frac{\Omega t}{2} & -\sin \frac{\Omega t}{2} \\ \sin \frac{\Omega t}{2} & \cos \frac{\Omega t}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (4.37b)$$

$$= \begin{bmatrix} \cos \frac{\Omega t}{2} \\ \sin \frac{\Omega t}{2} \end{bmatrix}. \quad (4.37c)$$

2 Interpretation

From Eq. 4.37 we see that the probabilities of finding the system in $|0\rangle$ and $|1\rangle$ at time t are given by

$$P_0(t) = \cos^2 \frac{\Omega t}{2}, \quad (4.38a)$$

$$P_1(t) = \sin^2 \frac{\Omega t}{2}. \quad (4.38b)$$

As depicted in Fig. 4.3, the system oscillates between $|0\rangle$ and $|1\rangle$ with a period of $2\pi/\Omega$. The key transitions are:

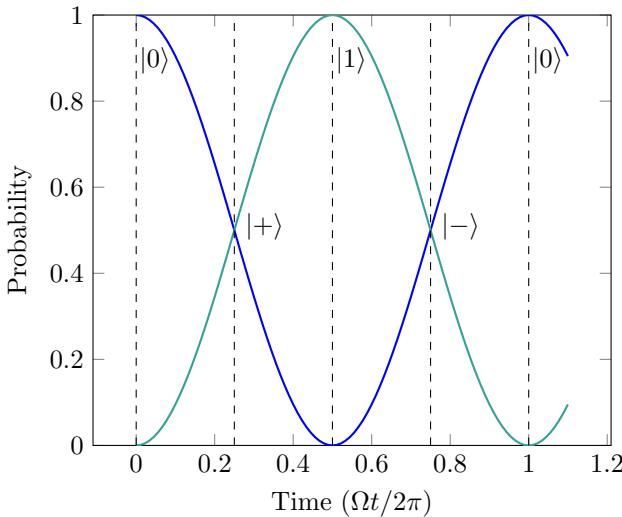


Figure 4.3: Plot of the probabilities $P_0(t)$ and $P_1(t)$ as functions of time.

- $t = 0$: system is in $|0\rangle$.
- $t = 0.25 \cdot 2\pi/\Omega$: system is in $|+\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.
- $t = 0.5 \cdot 2\pi/\Omega$: transitioned to $|1\rangle$.
- $t = 0.75 \cdot 2\pi/\Omega$: system is in $|-\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.
- $t = 1.0 \cdot 2\pi/\Omega$: back to $|0\rangle$.

3 Implementation of the Rotation Gates

Rabi oscillations, which involve the periodic exchange of population between quantum states, serve as a cornerstone for precise qubit control in quantum computing. They are essential for implementing various quantum gates and facilitating quantum information processing.

By applying a driving field for a time interval Δt , the qubit state undergoes a rotation by an angle $\gamma = \Omega\Delta t$ about the y -axis. This operation corresponds to the $R_y(\gamma)$ gate. In fact, the unitary operator $U(\Delta t)$ in Eq. 4.36 aligns exactly with the matrix representation of $R_y(\gamma)$ (refer to § 5.2.5 for details).

Through Rabi oscillations driven by Hamiltonians involving the X operator, the R_x gate can also be implemented. Other fundamental gates such as X , Y , H , S , and T can be realized by employing combinations of the R_x , R_y , and R_z gates, a topic to be discussed in Chapter 5.

Intuitively, Rabi oscillations can facilitate transformations of a qubit from the state $|0\rangle$ to $|1\rangle$ (analogous to an X gate operation) or to $|+\rangle$ (analogous to a Hadamard gate operation). Furthermore, this mechanism can be extended to coupled states between qubits, thereby enabling entanglement operations.

4 The Hamiltonian

The Hamiltonian for Rabi oscillations has a general form

$$H = \frac{\hbar\Omega}{2}(\cos\phi X + \sin\phi Y), \quad (4.39)$$

where ϕ is a phase factor that determines the specific combination of the Pauli matrices X and Y involved in the Hamiltonian. In the special case where $\phi = 0$, we have the X term; if $\phi = \pi/2$, we have the Y term. In the above discussion, for simplicity, we used only the Y term to illustrate the fundamental concepts.

The specific Hamiltonian for a real physical system depends on the direction of the external driving field, and its phase of oscillation.

Exercise 4.8 Given the Hamiltonian $H = \frac{\hbar\Omega}{2}X$, find a solution to the Schrödinger equation with initial condition $|\psi(0)\rangle = |1\rangle$. Calculate the probability $P_0(t) = |\langle 0|\psi(t)\rangle|^2$.

5 *The Rotating Wave Approximation

The Hamiltonian for Rabi oscillations, representing the energy of a two-level system in a high-frequency oscillatory field, can become quite complex due to the field's time-varying nature. The Rotating Wave Approximation (RWA) is a technique used in quantum mechanics to simplify this Hamiltonian, as described in [8]. It is particularly useful for analyzing interactions between a quantum system and an oscillating external field, such as laser or microwave radiation. The RWA is most valid near resonance, where the frequency of the external field closely matches the system's natural transition frequency, denoted by $\hbar\omega \approx E_1 - E_0$.

In the RWA, the Hamiltonian is transformed into an interaction representation, separating the free evolution of the system from its interaction with the driving field. The key step in RWA is to neglect the rapidly oscillating terms in the interaction Hamiltonian, which effectively average out to zero over time, focusing only on terms that oscillate at the difference frequency between the driving field and the natural frequency of the system.

This simplification yields an effective two-level Hamiltonian, as exemplified in Eq. 4.39, focusing primarily on resonant transitions. Within this framework, the effective Hamiltonian for Rabi oscillations can be viewed as a perturbation to the base Hamiltonian of the quantum system. This approximation greatly facilitates the analysis of quantum dynamics and phenomena, making the mathematical treatment more manageable, as demonstrated in our example. However, it's important to note that the RWA may not always be applicable, especially when far from resonance or when the terms omitted by the approximation become significant. As with any approximation in physics, its applicability must be carefully assessed within the specific physical context.

4.6 *Further Exploration

These topics are seldom covered in the context of quantum computing, except in specialized quantum simulations. They are included here for background information only. For further details, interested readers are directed to Ref. [8].

4.6.1 * Time Evolution Equations for Relativistic Systems

A common misconception is that quantum mechanics is synonymous with the Schrödinger equation. However, this equation is not universally applicable to all quantum systems. In the realm of relativistic systems, which involve high energies or particles moving at speeds close to the speed of light, different equations govern their time evolution.

For fermions, particles with half-integer spin, the Dirac equation is used to describe their evolution over time. The Dirac equation is given by:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H_D |\psi(t)\rangle, \quad (4.40)$$

where H_D is the Dirac Hamiltonian operator.

For bosons, particles with integer spin, the Klein-Gordon equation is used to describe their evolution over time. The Klein-Gordon equation is given by:

$$\square\phi + \frac{m^2 c^2}{\hbar^2}\phi = 0, \quad (4.41)$$

where $\square = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2$ is the d'Alembertian operator, a second-order differential operator defined in spacetime, m is the mass of the particle, c is the speed of light, and \hbar is the reduced Planck constant.

4.6.2 * Open Quantum Systems

In quantum mechanics, the Evolution Postulate and the Schrödinger equation are traditionally used to describe the dynamics of closed quantum systems, which are isolated from external influences or interactions. However, in the realm of quantum computing, qubits often undergo controlled interactions with external fields, such as microwave, lasers, or magnetic fields. This situation might appear to conflict with the closed-system framework, but it can be reconciled using perturbation theory.

In such scenarios, the external controls are typically treated as perturbations to the system's Hamiltonian. This approach allows for an analysis of the system's evolution under the influence of these controls. Indeed, Larmor precession and Rabi oscillations, as discussed in this chapter, are examples of how quantum computing often employs such controlled interactions. By modeling these interactions as perturbations, the foundational principles of quantum mechanics are extended to encompass the interaction of qubits with external control mechanisms. This extension is crucial for the precise control and manipulation of qubit states, forming a cornerstone of quantum information processing. It is this nuanced application of the Schrödinger equation and perturbation theory that bridges the gap between the theoretical model of closed-system dynamics and the interactive environment essential to quantum computation.

In quantum mechanics, open quantum systems that interact with their environment typically undergo non-unitary evolution. Such systems are comprehensively described by frameworks like quantum master equations or quantum trajectories, which account for time-dependent system-environment interactions. A notable instance of non-unitary evolution is the measurement process, where the state vector collapses to an eigenstate of the observable being measured. Additionally,

decoherence, a phenomenon where qubit states lose coherence due to uncontrolled environmental influences and noise, represents another form of non-unitary evolution. These non-unitary processes can be effectively modeled using Completely Positive Trace-Preserving (CPTP) maps, providing a mathematical framework essential in quantum information theory. We will delve into CPTP maps and quantum channels in § 12.2.7.

4.7 *Deferred Proofs

1 Hermitian Property of the Hamiltonian

As a result of the unitary evolution postulate (Eq. 4.1), the Hamiltonian operator H in the Schrödinger equation (Eq. 4.3) must be Hermitian. This can be seen as follows:

Consider an infinitesimal time interval Δt . Eq. 4.1 can be expressed as:

$$|\psi(\Delta t)\rangle - |\psi(0)\rangle = (U(\Delta t) - I)|\psi(0)\rangle. \quad (4.42)$$

At the same time, the Schrödinger equation Eq. 4.3 gives us

$$|\psi(\Delta t)\rangle - |\psi(0)\rangle = -i\frac{\Delta t}{\hbar}H|\psi(0)\rangle. \quad (4.43)$$

Comparing the two equations above, we obtain:

$$U(\Delta t) = I - i\frac{\Delta t}{\hbar}H. \quad (4.44)$$

Since $U(\Delta t)$ is unitary,

$$I = U(\Delta t)U(\Delta t)^\dagger = \left(I - i\frac{\Delta t}{\hbar}H\right)\left(I + i\frac{\Delta t}{\hbar}H^\dagger\right) = I - i\frac{\Delta t}{\hbar}H + i\frac{\Delta t}{\hbar}H^\dagger. \quad (4.45)$$

Here since Δt is infinitesimally small, we have ignored the $(\Delta t)^2$ term.

Apparently, for the above equation to hold, we must have $H = H^\dagger$, which indicates H is a Hermitian operator. Conversely, we can also show if H is Hermitian, the time evolution operator U as a solution to the Schrödinger equation is unitary.

2 Derivation of the General Solution to SE (Eq. 4.5)

Given the definition of e^A in Eq. 4.7, we can show that

$$\frac{\partial}{\partial t}e^{-\frac{i}{\hbar}Ht} = -\frac{i}{\hbar}He^{-\frac{i}{\hbar}Ht}. \quad (4.46)$$

Differentiate Eq. 4.5 over t :

$$\frac{\partial}{\partial t}|\psi(t)\rangle = \frac{\partial}{\partial t}\left(e^{-\frac{i}{\hbar}Ht}|\psi(0)\rangle\right) \quad (4.47a)$$

$$= -\frac{i}{\hbar}He^{-\frac{i}{\hbar}Ht}|\psi(0)\rangle \quad (4.47b)$$

$$= -\frac{i}{\hbar}H|\psi(t)\rangle. \quad (4.47c)$$

This is the same as the time-independent Schrödinger equation Eq. 4.4.

3 Exponential Product Formula (Eq. 4.8)

We will prove that

$$e^{A+B} = e^A e^B, \quad \text{if } [A, B] = 0. \quad (4.8)$$

If $[A, B] = 0$, meaning A and B commute, we can freely exchange the order of A and B in their product, $AB = BA$.

Given the definition of e^A in Eq. 4.7, we first consider the expansion of e^{A+B} :

$$e^{A+B} = \sum_{n=0}^{\infty} \frac{1}{n!} (A+B)^n. \quad (4.48)$$

Using the binomial theorem, $(A+B)^n$ can be expanded as:

$$(A+B)^n = \sum_{k=0}^n \binom{n}{k} A^k B^{n-k}. \quad (4.49)$$

Then

$$e^{A+B} = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} A^k B^{n-k} \quad (4.50a)$$

$$= \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{1}{k!(n-k)!} A^k B^{n-k}. \quad (4.50b)$$

Rearranging the sums, this becomes:

$$e^{A+B} = \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} \frac{1}{k!(n-k)!} A^k B^{n-k} \quad (4.51a)$$

$$= \left(\sum_{k=0}^{\infty} \frac{1}{k!} A^k \right) \left(\sum_{l=0}^{\infty} \frac{1}{l!} B^l \right) \quad (4.51b)$$

$$= e^A e^B. \quad (4.51c)$$

Thus, we have shown that $e^{A+B} = e^A e^B$ when A and B commute.

4 Derivation of Eq. 4.10

We want to prove that, if $A^2 = I$ and θ is a real number, then

$$e^{i\theta A} = \cos \theta I + i \sin \theta A. \quad (4.10)$$

Given the definition of the exponential of an operator,

$$e^{i\theta A} = \sum_{n=0}^{\infty} \frac{1}{n!} (i\theta A)^n, \quad (4.52)$$

we separate the series into even and odd terms:

$$e^{i\theta A} = \sum_{n=0}^{\infty} \frac{1}{(2n)!} (i\theta A)^{2n} + \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} (i\theta A)^{2n+1}. \quad (4.53)$$

Given $A^2 = I$, we can simplify the terms.

For even powers, $(i\theta A)^{2n} = (i^2)^n \theta^{2n} A^{2n} = (-1)^n \theta^{2n} I$.

For odd powers, $(i\theta A)^{2n+1} = (i\theta A)(-1)^n \theta^{2n} A^{2n} = i(-1)^n \theta^{2n+1} A$.

Thus, the series can be rewritten as:

$$e^{i\theta A} = \sum_{n=0}^{\infty} \frac{(-1)^n \theta^{2n}}{(2n)!} I + \sum_{n=0}^{\infty} \frac{i(-1)^n \theta^{2n+1}}{(2n+1)!} A \quad (4.54a)$$

$$= \left(\sum_{n=0}^{\infty} \frac{(-1)^n \theta^{2n}}{(2n)!} \right) I + i \left(\sum_{n=0}^{\infty} \frac{(-1)^n \theta^{2n+1}}{(2n+1)!} \right) A \quad (4.54b)$$

$$= \cos \theta I + i \sin \theta A. \quad (4.54c)$$

In the final step, we recognize the Taylor series expansions of $\cos \theta$ and $\sin \theta$. Therefore, we have proved that $e^{i\theta A} = \cos \theta I + i \sin \theta A$ when $A^2 = I$.

4.8 Summary and Conclusions

Temporal Evolution of Quantum States

In this chapter, we delved into the complex dynamics of quantum systems, centering on the key inquiry: How do qubit states evolve over time? The foundational principles of quantum evolution govern both the operation of quantum gates and the process of quantum annealing. To elucidate these principles further, we employed the case study of a single qubit as a representative example.

Core Principles in Quantum Mechanics

Anchoring our discussions was the Time Evolution Postulate, a cornerstone in quantum mechanics, which posits the deterministic unitary evolution of closed quantum systems. This provided the scaffolding to introduce the Schrödinger equation, the vital equation connecting the temporal evolution of quantum states with their energy via the Hamiltonian. The stationary nature of energy eigenstates, a direct implication of the time-independent Schrödinger equation, was also delineated, elucidating the distinction between ground states and excited states and their pertinence in quantum computing.

Quantum Computing Paradigms

The realm of quantum computing has witnessed the emergence of two cardinal paradigms: Gate-based universal quantum computing and adiabatic quantum computing, including quantum annealing. Their distinctions and foundational underpinnings were examined, underscoring their unique methodologies and applications, all rooted in the principle of unitary evolution of quantum states.

Insights into Quantum Systems

Larmor Precession and Rabi Oscillations, two quintessential phenomena in quantum mechanics, were explored to furnish insights into the behavior of quantum systems

under the influence of external fields. These served as pedagogical tools, illuminating essential techniques for tackling the Schrödinger equation, particularly relevant to the domain of quantum hardware.

Perspective on Quantum Systems

The chapter culminated with an exploration into topics not directly applied in quantum computing, yet important for a holistic understanding of the subject. These encompassed relativistic time evolution equations and the domain of open quantum systems.

Upcoming Topics

In the next chapter, we will explore quantum gates, which are central to quantum computing. Building on our foundational understanding of quantum mechanics, we will study the basic single-qubit gates before advancing to topics like multi-qubit gates, entanglement, quantum teleportation, and the E91 quantum key distribution protocol. These discussions will connect our theoretical understanding of quantum mechanics with practical quantum computing applications.

Problem Set 4

- 4.1** Consider a one-qubit quantum gate defined by the Hamiltonian $H = \hbar\omega X$, where X is the Pauli- X operator. Derive the unitary evolution operator $U(\Delta t)$ over an interval Δt and elucidate the influence of this gate on an arbitrary qubit state.

- 4.2** Consider $U = u_x X + u_y Y + u_z Z$, where X, Y, Z are the Pauli operators, and u_x, u_y, u_z are real numbers with $u_x^2 + u_y^2 + u_z^2 = 1$. Express $e^{i\gamma U}$ as a single 2×2 matrix, where γ is a real number.

Hint: Show that $U^2 = I$ first.

- 4.3** Express e^X as a single 2×2 matrix, where X is the Pauli- X operator. (Note this is not the same as e^{iX} .)

- 4.4** Find a 2×2 Hamiltonian H such that e^{-iH} equals the Hadamard matrix $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

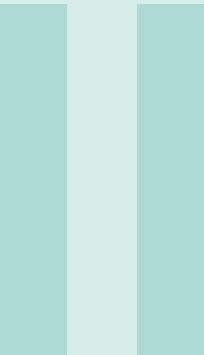
- 4.5** Given the Hamiltonian $H = \hbar\omega \begin{bmatrix} 0 & \alpha \\ \alpha & 1 \end{bmatrix}$, where $\alpha, \omega \in \mathbb{R}$, find the energy levels E_0 and E_1 , and corresponding eigenstates $|E_0\rangle$ and $|E_1\rangle$.

- 4.6** Consider a Hamiltonian $H = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$.

(a) Diagonalize this Hamiltonian to find its eigenvalues and eigenvectors.

(b) Given the time-evolution operator $U(t) = e^{-iHt}$, compute $U(t)$ for a time t , using the eigendecomposition of H .

- (c) If we start with a state $|0\rangle$, and we evolve it using $U(t)$, the resulting state will be $U(t)|0\rangle$. Compute the probabilities of measuring $|0\rangle$ and $|1\rangle$ after this evolution.
- 4.7** (Rabi oscillation with general Hamiltonian.) Consider the Hamiltonian $H = \frac{\hbar\Omega}{2}(\cos\phi X + \sin\phi Y)$, where $\phi \in \mathbb{R}$.
- Find the matrix form of $U(t) = e^{-iHt/\hbar}$.
 - Given the initial condition $|\psi(0)\rangle = |0\rangle$, calculate the probability evolution $P_0(t) = |\langle 0|\psi(t)\rangle|^2$.
- 4.8** Consider a Hamiltonian $H = \frac{T-t}{T}Z + \frac{t}{T}X$, where X and Z are the Pauli matrices, and T is the total evolution time (i.e., $t \in [0, T]$). Write an approximate solution to the Schrödinger equation using Trotterization.
- 4.9** Find an $\alpha \in \mathbb{R}$ such that $e^{i\alpha Z}Xe^{-i\alpha Z} = Y$, where X, Y, Z are the Pauli matrices. Interpret the meaning of this relation in terms of qubit-state rotations and quantum gates.



Quantum Gates & Elementary Circuits

5	Single-Qubit Quantum Gates	107
6	Multi-Qubit Systems	138
7	Multi-Qubit Quantum Gates	170

5. Single-Qubit Quantum Gates

Contents

5.1	Quantum Versus Classical Logic Gates	108
5.1.1	Quantum Gates as Unitary Operators	109
5.1.2	Reversibility Requirement	109
5.1.3	No-Cloning Requirement	110
5.1.4	Linearity	110
5.1.5	From Quantum Gates to Circuits	111
5.2	Common Single-Qubit Gates	111
5.2.1	Approaches of Analysis: X Gate Example	111
5.2.2	Pauli Gates	113
5.2.3	The Hadamard H Gate	114
5.2.4	The Phase Gates	115
5.2.5	The R_x , R_y , and R_z Rotation Gates	116
5.2.6	*The Unified Rotation Gate	118
5.3	From Gate Sequences to Quantum Circuits	119
5.3.1	The Circuit Model of Quantum Computing	119
5.3.2	Quantum Circuit Diagrams	120
5.3.3	Gate Sequences	121
5.4	Quantum Random Number Generator	122
5.5	The BB84 Quantum Key Distribution (QKD) Protocol	124
5.5.1	Fundamentals of Cryptography	124
5.5.2	Introduction to the BB84 QKD Protocol	125
5.5.3	Procedure of the BB84 QKD Protocol	126
5.5.4	Detection of Eavesdropping	127
5.5.5	BB84 QKD Implementation with Qubits	128
5.5.6	*Further Exploration	130
5.6	The Quantum Coin Game	131
5.6.1	The Game Rules	131
5.6.2	Alice Always Wins!	132
5.6.3	Bob's Strategy Shift	133
5.6.4	Alice Overcomes Bob's New Gambit	133

5.7	*The No-Cloning Theorem: Proof Outline	134
5.8	Summary and Conclusions	135
	Problem Set 5	136

This chapter marks the beginning of our exploration into the world of quantum gates, the critical building blocks of quantum computing systems.

The principles of quantum computing are deeply intertwined with the fundamental tenets of quantum mechanics. The relationship between the two can be likened to how Newtonian mechanics underpins the launching of satellites. Both fields require an in-depth understanding of the respective theoretical principles and meticulous engineering to ensure precise real-world application.

Just as launching a satellite involves grasping Newtonian principles like gravity and motion to craft the right trajectory and propulsion systems, creating quantum gates and circuits requires understanding the key principles of quantum mechanics, such as superposition and entanglement. This knowledge enables us to control qubits and construct quantum algorithms for tackling complex computational problems. In both contexts, we leverage basic physical theories to devise practical solutions.

In previous discussions, we have laid down the foundational principles of quantum mechanics—including quantum states, measurements, and unitary evolution—and we have developed a solid understanding of qubits. Now, we’re ready to extend our exploration into the fascinating world of quantum gates.

We’ll commence our exploration with single-qubit gates and their various applications. This groundwork will then pave the way for more advanced topics, including multi-qubit gates, entanglement, quantum teleportation, and the E91 quantum key distribution protocol.

5.1 Quantum Versus Classical Logic Gates

Just as classical logic gates are fundamental to classical computing, quantum gates form the cornerstone of quantum computing. Both types of gates serve as standardized and highly optimized components within their respective computational paradigms. Quantum gates are specifically designed to manipulate quantum states, thereby facilitating the execution of quantum algorithms. These gates can be viewed as specialized solution cases of the Schrödinger equation, tailored for quantum computational tasks. Although implemented differently across various qubit platforms, the operational functionality of quantum gates has remained consistent. This mirrors the evolution of classical gates, which have transitioned from vacuum tubes to transistors and other technologies, yet their core principles of simplified state representations and logical operations have remained unchanged.

5.1.1 Quantum Gates as Unitary Operators

Quantum gates introduce a distinct capability not found in classical computing: the ability to operate on superpositions of states. This unique feature arises from the principles of quantum mechanics that govern qubit behavior. The time evolution of a closed quantum system, as described by the Schrödinger equation, is characterized by unitary operators. Consequently, each quantum gate represents a unitary transformation, capable of acting on superpositions of quantum states and producing superpositions of outcomes.

As explored in § 4.4.1, the action of a quantum gate is analogous to the application of a specially designed Hamiltonian H over a specified time interval Δt . The physical implementation of a quantum gate, thus, depends on the precise engineering of this Hamiltonian.

The evolution of a quantum system's state, in accordance with the Schrödinger equation, is generally expressed as:

$$|\psi(\Delta t)\rangle = U(\Delta t) |\psi(0)\rangle, \quad (5.1)$$

where $U(\Delta t)$ is the unitary operator that describes the gate operation.

In quantum gate and circuit discussions, it's common to omit the Δt term. Thus, Eq. 5.1 is often represented as:

$$|\psi_1\rangle = U |\psi_0\rangle. \quad (5.2)$$

Here, $|\psi_0\rangle \equiv |\psi(0)\rangle$ denotes the system's state prior to the gate operation, $|\psi_1\rangle \equiv |\psi(\Delta t)\rangle$ symbolizes the state post-operation, and $U \equiv U(\Delta t)$ characterizes the operator signifying the gate's action.

5.1.2 Reversibility Requirement

One of the key attributes of quantum gates, stemming from their nature as unitary operators, is reversibility. Specifically, unitary operators possess a notable property: the inverse of a unitary operator is its conjugate transpose, denoted as $U^{-1} = U^\dagger$. Consequently, by applying the inverse operation U^{-1} , which is also unitary, one can effectively reverse the original quantum operation.

This aspect of quantum gates contrasts sharply with classical logic gates. Classical gates are not inherently reversible, as classical mechanics does not enforce the reversibility requirement inherent to quantum mechanics. Because of this not all classical gates have quantum counterparts.

For single-bit classical gates, we can identify four possible ones:

1. Identity (or “do nothing”): 0 maps to 0 and 1 maps to 1.
2. NOT (or bit-flip): 0 maps to 1 and 1 maps to 0.
3. Constant 0: Both 0 and 1 map to 0.
4. Constant 1: Both 0 and 1 map to 1.

It's clear that the third and fourth gates are non-reversible, as they both map two possible inputs to a single output. In quantum mechanics, these gates would violate the principle of unitarity.

Additionally, some well-known multi-bit classical gates are irreversible and hence, cannot be directly implemented as quantum gates:

- AND Gate: This gate has two input bits and one output bit. An output of 0 could arise from inputs 00, 01, or 10. Only the output 1 uniquely corresponds to the input 11. (See Fig. 5.1.)
- OR Gate: Analogous to the AND gate, the OR gate's output of 1 can originate from inputs 01, 10, or 11. An output of 0 is unique to 00.
- Other Gates such as NAND, NOR, and XOR: These gates also exhibit irreversibility for certain outputs.

While these classical gates can't be directly translated to quantum gates, alternative methods exist to emulate their functionality in quantum computing. One prevalent strategy involves using extra output bits (often termed as ancilla bits) to retain all input-bit information. This topic will receive further exploration in subsequent chapters.

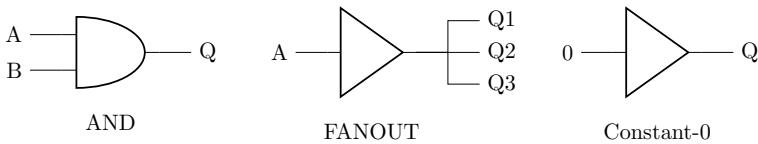


Figure 5.1: Examples of Classical Irreversible Gates

5.1.3 No-Cloning Requirement

Owing to the unitary nature of quantum evolution, there exists a pivotal principle in quantum computing and quantum information: the no-cloning theorem:

The No-Cloning Theorem.

It is fundamentally impossible to create an exact replica of an arbitrary, unknown quantum state solely using any physically realizable quantum operation while retaining the original state.

This theorem implies that one cannot perfectly clone or duplicate a qubit without altering its original state. As a consequence, operations like FANOUT (see Fig. 5.1), which are pervasive in classical digital circuits, are not feasible in quantum computing. On a positive note, the no-cloning principle underscores the robust security in quantum cryptography, exemplified by protocols such as the BB84.

A proof of this theorem is outlined at the end of this chapter (§ 5.7).

5.1.4 Linearity

Quantum gates, represented by complex unitary matrices, act linearly on quantum states. This linearity property, pivotal to quantum mechanics, ensures that any linear combination (superposition) of states is again a quantum state. It allows quantum gates to operate on superpositions, resulting in quantum parallelism and entanglement, which are among the core strengths of quantum computing.

5.1.5 From Quantum Gates to Circuits

Quantum circuits are constructed by composing sequences of quantum gates to implement quantum algorithms. The inherent nature of these circuits is unitary, provided they consist exclusively of quantum gates. This composition guarantees that the resulting operation remains unitary, thereby preserving the quantum state's norm and coherence within a closed system.

However, the scope of quantum circuits extends beyond solely unitary gates. Non-unitary operations, such as measurements and state resets, play a crucial role in quantum computing. While unitary gates facilitate the evolution of quantum states within a closed system, non-unitary operations enable interaction with the external environment. This leads to phenomena such as state collapse or decoherence. Such operations are indispensable for extracting information from quantum systems and are integral to certain steps in quantum algorithms, thereby introducing additional complexity and capabilities to quantum circuits that surpass those of classical logic gates.

The synergy of unitary and non-unitary operations within quantum circuits captures the full breadth of quantum computing capabilities. These range from maintaining coherent superpositions to executing measurements that provide classical information. As we progress in this chapter and in subsequent ones, we will delve into the intricate process of constructing quantum circuits and examine how they form the foundation of complex quantum algorithms.

5.2 Common Single-Qubit Gates

Now we are ready to turn to the common quantum gates involving a single qubit. As shown in Table 5.1, these include the Pauli- X , Y , Z gates, the Hadamard (H) gate, the phase gates, and the rotation gates. By combining these gates in different ways, quantum circuits can solve complex problems and perform tasks that are infeasible for classical computers, such as factoring large numbers or simulating quantum systems.

5.2.1 Approaches of Analysis: X Gate Example

The action of a qubit gate can be represented and analyzed in various ways, which are frequently used in different applications. It's essential to be familiar with all of them. As an illustrative example, we will explore the X gate.

The X gate, also known as the Pauli- X gate, is a fundamental quantum gate, acting as the quantum counterpart to the classical NOT gate. It maps $|0\rangle$ to $|1\rangle$ and vice versa. Therefore, it is also known as the bit-flip gate.

1 Matrix Representation

The X gate is represented by the Pauli- X matrix:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (5.3)$$

Given the column vectors

Gate	Symbol	Operator	Description
Identity (I)	$\begin{array}{ c } \hline I \\ \hline \end{array}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	Leaves the qubit unchanged.
Pauli- X (X)	$\begin{array}{ c } \hline X \\ \hline \end{array}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	NOT or bit-flip gate, maps $ 0\rangle$ to $ 1\rangle$ and $ 1\rangle$ to $ 0\rangle$.
Pauli- Z (Z)	$\begin{array}{ c } \hline Z \\ \hline \end{array}$	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	Phase-flip gate, maps $ 0\rangle$ to $ 0\rangle$ and $ 1\rangle$ to $- 1\rangle$.
Pauli- Y (Y)	$\begin{array}{ c } \hline Y \\ \hline \end{array}$	$\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$	Combines bit flip and phase flip operations.
Hadamard (H)	$\begin{array}{ c } \hline H \\ \hline \end{array}$	$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$	Creates superposition. Maps $ 0\rangle$ to $ +\rangle$ and $ 1\rangle$ to $ -\rangle$.
S (\sqrt{Z})	$\begin{array}{ c } \hline S \\ \hline \end{array}$	$\begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$	Applies $\frac{\pi}{2}$ phase shift to the $ 1\rangle$ state.
T ($\sqrt[4]{Z}$)	$\begin{array}{ c } \hline T \\ \hline \end{array}$	$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{bmatrix}$	Applies $\frac{\pi}{4}$ phase shift to the $ 1\rangle$ state.
Phase ($P(\theta)$)	$\begin{array}{ c } \hline P(\theta) \\ \hline \end{array}$	$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix}$	Applies θ phase shift to the $ 1\rangle$ state.
$R_x(\theta)$	$\begin{array}{ c } \hline R_x(\theta) \\ \hline \end{array}$	$\begin{bmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix}$	Rotates qubit around x -axis of the Bloch sphere by θ .
$R_y(\theta)$	$\begin{array}{ c } \hline R_y(\theta) \\ \hline \end{array}$	$\begin{bmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix}$	Rotates qubit around y -axis of the Bloch sphere by θ .
$R_z(\theta)$	$\begin{array}{ c } \hline R_z(\theta) \\ \hline \end{array}$	$\begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix}$	Rotates qubit around z -axis of the Bloch sphere by θ .

Table 5.1: Common Single-Qubit Quantum Gates

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad (5.4)$$

the bit-flip transformation becomes evident: $X|0\rangle = |1\rangle$ and $X|1\rangle = |0\rangle$.

2 Bra-Ket Representation

A matrix U can be expanded using the outer products of the computational basis states:

$$U = \begin{bmatrix} u_{00} & u_{01} \\ u_{10} & u_{11} \end{bmatrix} = u_{00}|0\rangle\langle 0| + u_{01}|0\rangle\langle 1| + u_{10}|1\rangle\langle 0| + u_{11}|1\rangle\langle 1|. \quad (5.5)$$

Using this expansion, the X gate matrix can be represented in the bra-ket notation:

$$X = |0\rangle\langle 1| + |1\rangle\langle 0|. \quad (5.6)$$

Exercise 5.1 Verify $X|0\rangle = |1\rangle$ and $X|1\rangle = |0\rangle$ using bra-ket notation.

3 Applying to Superposition States

Quantum gates, being linear operators, can operate on superposition states. When the X gate acts on

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad (5.7)$$

the resulting state becomes

$$X|\psi\rangle = \beta|0\rangle + \alpha|1\rangle = \begin{bmatrix} \beta \\ \alpha \end{bmatrix}. \quad (5.8)$$

Exercise 5.2 Verify Eq. 5.8 using both the matrix and bra-ket representations of X .

4 Native in $\{|+\rangle, |-\rangle\}$ Basis

The Pauli matrices are both unitary and Hermitian. As such, X possesses eigenvectors that form the $\{|+\rangle, |-\rangle\}$ basis (see also § 3.3.3):

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad (5.9a)$$

$$|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \quad (5.9b)$$

Applying the X gate to these vectors yields:

$$X|+\rangle = |+\rangle, \quad X|-\rangle = -|-\rangle. \quad (5.10)$$

5 Rotation on Bloch Sphere

In the Bloch sphere representation, the X gate corresponds to a 180° rotation about the x -axis. As shown in Fig. 2.4, the state vector $|\psi\rangle$ moves 180° along the dotted circle centered around the x -axis.

6 Boolean Representation

In addition to matrix and bra-ket representations, many quantum gates can also be expressed using Boolean representations. We will explore this in detail in § 7.3.

5.2.2 Pauli Gates

We have just delved into a detailed examination of the X gate, introducing various analytical tools in the process. These tools can be extended to the entire family of Pauli gates, which comprise the X , Y , Z , and I gates.

The Z gate, known as the phase-flip gate, imparts a π phase to the $|1\rangle$ state. Additionally, it serves as the flip gate for the $|+\rangle$ and $|-\rangle$ states, transitioning $|+\rangle$ to $|-\rangle$ and vice versa.

The Y gate integrates both the bit flip and phase flip. Interestingly, it can be shown that $ZX = iY$, a derivation we invite the reader to explore in the exercises below.

Gate	Bra-Ket Operators	$ 0\rangle, 1\rangle \rightarrow$	$ +\rangle, -\rangle \rightarrow$	$\alpha 0\rangle + \beta 1\rangle \rightarrow$
I	$ 0\rangle\langle 0 + 1\rangle\langle 1 $	$I 0\rangle = 0\rangle$	$I +\rangle = +\rangle$	$\alpha 0\rangle + \beta 1\rangle$
	$ +\rangle\langle + + -\rangle\langle - $	$I 1\rangle = 1\rangle$	$I -\rangle = -\rangle$	
X	$ 1\rangle\langle 0 + 0\rangle\langle 1 $	$X 0\rangle = 1\rangle$	$X +\rangle = +\rangle$	
	$ +\rangle\langle + - -\rangle\langle - $	$X 1\rangle = 0\rangle$	$X -\rangle = -\rangle$	$\alpha 1\rangle + \beta 0\rangle$
Z	$ 0\rangle\langle 0 - 1\rangle\langle 1 $	$Z 0\rangle = 0\rangle$	$Z +\rangle = -\rangle$	
	$ -\rangle\langle + + +\rangle\langle - $	$Z 1\rangle = - 1\rangle$	$Z -\rangle = +\rangle$	$\alpha 0\rangle - \beta 1\rangle$
Y	$i(1\rangle\langle 0 - 0\rangle\langle 1)$	$Y 0\rangle = i 1\rangle$	$Y +\rangle = -i -\rangle$	
	$i(+\rangle\langle - - -\rangle\langle +)$	$Y 1\rangle = -i 0\rangle$	$Y -\rangle = i +\rangle$	$i\alpha 1\rangle - i\beta 0\rangle$
H	$ +\rangle\langle 0 + -\rangle\langle 1 $	$H 0\rangle = +\rangle$	$H +\rangle = 0\rangle$	
	$ 0\rangle\langle + + 1\rangle\langle - $	$H 1\rangle = -\rangle$	$H -\rangle = 1\rangle$	$\alpha +\rangle + \beta -\rangle$
S	$ 0\rangle\langle 0 + i 1\rangle\langle 1 $	$S 0\rangle = 0\rangle$ $S 1\rangle = i 1\rangle$	-	$\alpha 0\rangle + i\beta 1\rangle$
T	$ 0\rangle\langle 0 + e^{i\frac{\pi}{4}} 1\rangle\langle 1 $	$T 0\rangle = 0\rangle$ $T 1\rangle = e^{i\frac{\pi}{4}} 1\rangle$	-	$\alpha 0\rangle + e^{i\frac{\pi}{4}}\beta 1\rangle$
P	$ 0\rangle\langle 0 + e^{i\theta} 1\rangle\langle 1 $	$P(\theta) 0\rangle = 0\rangle$ $P(\theta) 1\rangle = e^{i\theta} 1\rangle$	-	$\alpha 0\rangle + e^{i\theta}\beta 1\rangle$

Table 5.2: Transformations by Common Single-Qubit Quantum Gates

The I gate stands for the identity gate, represented by the identity matrix. While it has no transformative effect on the qubit state and doesn't require physical implementation, its frequent use in circuit analysis justifies its inclusion for completeness.

All Pauli gates are involutory, meaning applying a Pauli gate twice restores the qubit state to its original form. For example, $X^2 = I$.

For quick reference, essential formulas related to these gates and others we will explore subsequently can be found in Tables 5.1 and 5.2. While we won't delve into all the formulas here, we encourage readers to tackle the exercises provided below.

Exercise 5.3 Show that Z acts as the flip gate for $|+\rangle$ and $|-\rangle$, meaning $Z|+\rangle = |-\rangle$ and $Z|-\rangle = |+\rangle$, using both matrix and bra-ket representations.

Exercise 5.4 Show that $Y(\alpha|0\rangle + \beta|1\rangle) = i(\alpha|1\rangle - \beta|0\rangle)$ employing both matrix and bra-ket representations.

5.2.3 The Hadamard H Gate

The Hadamard (H) gate is fundamental in quantum computing. It is commonly used to create superposition states as it transforms $|0\rangle$ to $|+\rangle$, and $|1\rangle$ to $|-\rangle$:

$$H |0\rangle = |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad (5.11a)$$

$$H |1\rangle = |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \quad (5.11b)$$

For this reason, the $\{|+\rangle, |-\rangle\}$ basis is often referred to as the Hadamard basis.

Being involutory, $H^2 = I$. This means applying H twice restores the qubit state to its original form:

$$H |+\rangle = |0\rangle, \quad H |-\rangle = |1\rangle. \quad (5.12)$$

Lastly,

$$H(\alpha |0\rangle + \beta |1\rangle) = \alpha |+\rangle + \beta |-\rangle. \quad (5.13)$$

Exercise 5.5 Show that $H |+\rangle = |0\rangle$ and $H |-\rangle = |1\rangle$ using both matrix and bra-ket representations.

5.2.4 The Phase Gates

Phase gates are instrumental in introducing a relative phase between the basis states of a qubit. Among these gates, the most commonly utilized are the S , T , and P gates.

The S gate, often termed the “phase gate,” adds a phase of $\pi/2$ (or 90°) to the $|1\rangle$ state. It can also be represented as the square root of the Z gate:

$$S |0\rangle = |0\rangle, \quad S |1\rangle = i |1\rangle, \quad S = \sqrt{Z}. \quad (5.14)$$

The T gate introduces a more nuanced phase of $\pi/4$ (or 45°). It is effectively the fourth root of the Z gate:

$$T |0\rangle = |0\rangle, \quad T |1\rangle = e^{i\pi/4} |1\rangle, \quad T = \sqrt[4]{Z}. \quad (5.15)$$

The general phase gate, $P(\phi)$, allows for a custom phase ϕ to be added to the $|1\rangle$ state:

$$P(\phi) |0\rangle = |0\rangle, \quad P(\phi) |1\rangle = e^{i\phi} |1\rangle. \quad (5.16)$$

These gates are foundational for tasks such as phase kickback and quantum Fourier transform. By adjusting the relative phase, they offer an extra degree of freedom and control in quantum algorithms.

Exercise 5.6 Show that $S^2 = Z$ and $T^4 = Z$.

Exercise 5.7 Demonstrate how the S , T , and $P(\phi)$ gates transform the general qubit state $\alpha |0\rangle + \beta |1\rangle$ using the bra-ket notation.

5.2.5 The R_x , R_y , and R_z Rotation Gates

Rotation gates play a crucial role in quantum computing, offering a method to rotate qubit states around various axes of the Bloch sphere (see Fig. 5.2). The three primary rotation gates are R_x , R_y , and R_z , corresponding to rotations about the x , y , and z axes, respectively.

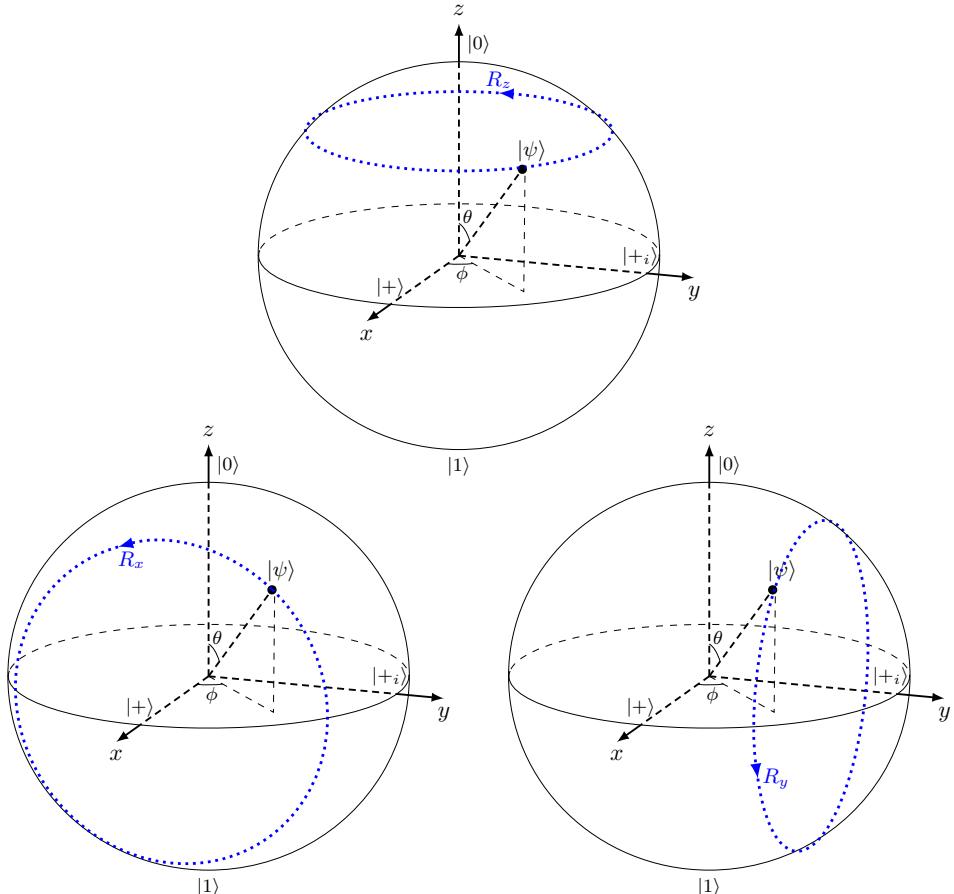


Figure 5.2: R_x , R_y , and R_z Gates as Rotations on the Bloch Sphere

1. The $R_x(\theta)$ gate rotates a qubit state around the x -axis by an angle θ :

$$R_x(\theta) = e^{-i\frac{\theta}{2}X} = \begin{bmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix}. \quad (5.17)$$

Exercise 5.8 Demonstrate that the matrix of $R_x(\theta)$ is unitary. Is it also Hermitian?

2. Similarly, the $R_y(\theta)$ gate rotates a qubit state around the y -axis:

$$R_y(\theta) = e^{-i\frac{\theta}{2}Y} = \begin{bmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix}. \quad (5.18)$$

3. The $R_z(\theta)$ gate performs a rotation around the z -axis:

$$R_z(\theta) = e^{-i\frac{\theta}{2}Z} = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix}. \quad (5.19)$$

A key property of these rotation operators is:

$$R_i(\theta)^\alpha = R_i(\alpha\theta), \quad (5.20)$$

where $\alpha \in \mathbb{R}$ and $i \in \{x, y, z\}$.

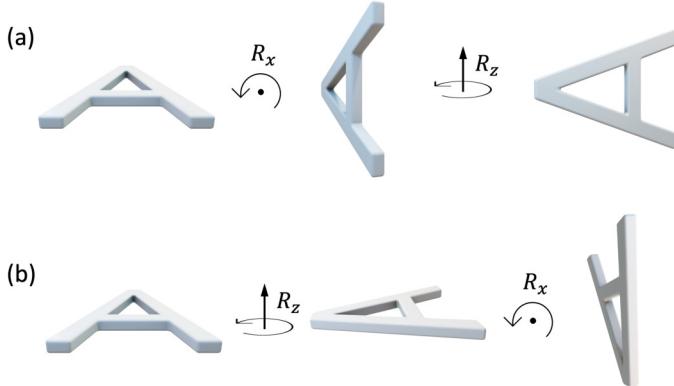


Figure 5.3: Demonstration of Non-Commutative Nature of 3D Rotations

It is imperative to understand that rotations in three-dimensional space are non-commutative when applied in different sequences. As depicted in Fig. 5.3, rotating around the x -axis followed by the z -axis (part a) does not result in the same orientation as when the rotation sequence is reversed (part b). This can be mathematically expressed as $R_z R_x \neq R_x R_z$, or equivalently $[R_z, R_x] \neq 0$.



The equations for R_x , R_y , and R_z closely resemble the solutions for spin precession (Eq. 4.26) and Rabi oscillations (Eq. 4.36). Hence, both phenomena serve as effective means to implement these gates.

The Pauli gates, Hadamard gate, and Phase gates can all be expressed as special cases of rotation gates, albeit with a potential global phase difference. The relationships between them are given by:

$$X = iR_x(\pi), \quad (5.21a)$$

$$Y = iR_y(\pi), \quad (5.21b)$$

$$Z = -R_z(\pi), \quad (5.21c)$$

$$S = e^{i\frac{\pi}{4}} R_z\left(\frac{\pi}{2}\right), \quad (5.21d)$$

$$T = e^{i\frac{\pi}{8}} R_z\left(\frac{\pi}{4}\right), \quad (5.21e)$$

$$H = iR_x(\pi)R_y\left(\frac{\pi}{2}\right) = iR_y\left(\frac{\pi}{2}\right)R_z(\pi). \quad (5.21f)$$

Equation 5.21 offers insight into understanding single-qubit gates as rotations on the Bloch sphere. For instance, the Hadamard gate can be interpreted as a

rotation of 90° around the y -axis, followed by a 180° rotation around the x -axis. Alternatively, it can also be seen as a 180° rotation around the z -axis, followed by a 90° rotation around the y -axis.

Exercise 5.9 Derive the expression for the X gate as a special case of the R_x rotation gate.

Exercise 5.10 Among the gates listed in Table 5.1, identify which gates are Hermitian (i.e., $U = U^\dagger$), which are involutory (i.e., $U^2 = I$), and which satisfy both conditions.

5.2.6 *The Unified Rotation Gate

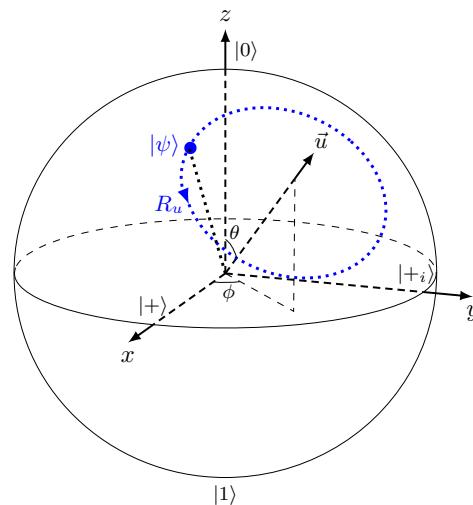


Figure 5.4: The Unified Rotation Gate (R_u) on the Bloch Sphere

We can further unify R_x , R_y , and R_z into a single rotation operator R_u . Below is how it's defined (refer to Fig. 5.4).

Given any real unit vector in three-dimensional space represented by $\hat{u} = (u_x, u_y, u_z)$, where the condition $u_x^2 + u_y^2 + u_z^2 = 1$ holds, we can define the unitary operator

$$\sigma_u = u_x X + u_y Y + u_z Z. \quad (5.22)$$

In the context of spherical coordinates with polar angle θ and azimuthal angle ϕ , the components are given by

$$u_x = \sin \theta \cos \phi, \quad u_y = \sin \theta \sin \phi, \quad u_z = \cos \theta. \quad (5.23)$$

This results in:

$$\sigma_u = \sin \theta \cos \phi X + \sin \theta \sin \phi Y + \cos \theta Z \quad (5.24a)$$

$$= \begin{bmatrix} \cos \theta & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta \end{bmatrix}. \quad (5.24b)$$



In this context, we use γ to denote the rotation angle, reserving θ for the polar directional angle.

For a rotation about the axis \hat{u} by an angle γ :

$$R_u(\gamma, \theta, \phi) = e^{-i\frac{\gamma}{2}\sigma_u} \quad (5.25a)$$

$$= \cos \frac{\gamma}{2} I - i \sin \frac{\gamma}{2} \sigma_u \quad (5.25b)$$

$$= \begin{bmatrix} \cos \frac{\gamma}{2} - i \sin \frac{\gamma}{2} \cos \theta & -i \sin \frac{\gamma}{2} \sin \theta e^{-i\phi} \\ -i \sin \frac{\gamma}{2} \sin \theta e^{i\phi} & \cos \frac{\gamma}{2} + i \sin \frac{\gamma}{2} \cos \theta \end{bmatrix}. \quad (5.25c)$$

It might be evident to the astute reader that all single-qubit gates can be represented as a special case of R_u . Then, why not just use R_u and eliminate all others? Primarily because R_u is parameterized by three values (θ , ϕ , and γ) and is challenging to implement physically.

Exercise 5.11 Express R_x , R_y , and R_z in terms of $R_u(\gamma, \theta, \phi)$ and verify that their matrix representations are equivalent.

5.3 From Gate Sequences to Quantum Circuits

Quantum gates, foundational to quantum circuits, dictate the progression of quantum computations on qubits. These circuits visually depict quantum algorithms, offering a clear understanding of their design and sequence. As we've delved into single-qubit gates, our next logical step is to explore sequences of these foundational gates within quantum circuits.

5.3.1 The Circuit Model of Quantum Computing

The circuit model of quantum computing emerges as a leading framework for understanding and implementing quantum algorithms and operations. Drawing parallels with classical circuits, which process information using a series of logical gates, quantum circuits use quantum gates to manipulate qubits. These quantum computations can be visually represented through circuit diagrams, delineating each gate and its sequence, thereby facilitating an intuitive grasp and design of quantum algorithms.

This model is tailored for universal quantum computing, enabling it to theoretically execute any quantum algorithm. However, it's important to recognize that the circuit model isn't the only quantum computational model out there. The quantum Turing machine model serves as another key model, offering a more abstract perspective of quantum computation. While rich in theoretical insights, its high-level nature tends to make it less suited for practical applications compared to the circuit model.

Diverging from these models is the concept of quantum annealing. This paradigm, exemplified by systems like D-Wave, excels in rapidly searching for optimized solutions within vast solution spaces, making it particularly effective for certain optimization problems. However, quantum annealing has its own set of limitations and does not replace the capabilities of universal quantum computers.

A comprehensive study of quantum computing, such as ours, naturally commences with the circuit model for universal computing due to its practical relevance, visual clarity, and wide applicability. This foundational understanding paves the way for diving deeper into other quantum computational paradigms and appreciating their unique strengths and challenges.

5.3.2 Quantum Circuit Diagrams

Quantum circuits are graphically depicted through circuit diagrams. For illustration, Fig. 5.5 will be used to elucidate some salient aspects of quantum circuit diagrams.

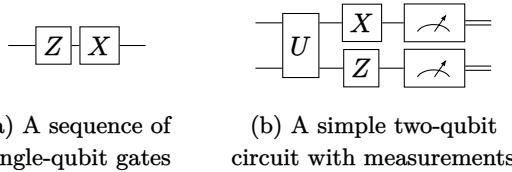


Figure 5.5: Examples of Quantum Circuits

1 Wires

Often referred to as “wires,” the horizontal lines depict the temporal evolution of individual qubits in the circuit. Each wire denotes a single qubit and charts its state alteration through various gate operations.

2 Gates

Symbols, frequently housed within rectangles or squares, represent the gates and are set upon the wires of the qubits they influence. These symbols signify quantum operations responsible for the metamorphosis of qubit states.

3 Left-to-right Sequence

Gates are arranged from left to right on the wires, encapsulating the chronological sequence of operations. Progressing from left to right mirrors the advancement in time, with the sequence of gates being pivotal, orchestrating the comprehensive state transformation of the qubits.

4 Order of Operator Multiplication

The composition of a sequence of gates is represented as the matrix product of the related gate matrices. Owing to conventional mathematical notation, the sequence of these matrices is in inverse order relative to the sequence of the gates. For instance, in Fig. 5.5 (a), the progression from a Z gate to an X gate results in the matrix product XZ .

5 Input and Output

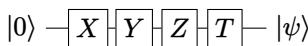
Positioned on the left of the circuit diagram, the initial states (often initialized to $|0\rangle$) are one form of input to the circuit. However, the input is not solely limited to such initializations; it also encompasses other elements such as query oracles and gate parametrization (to be discussed in Chapter 11). Conversely, the final states on the right side of the circuit diagram represent the output after all quantum gates

and operations have been applied. Measurement operations, typically represented by rectangles enclosing a meter sign, are introduced at this stage to extract classical data—indicated by double wires—from the quantum states of the qubits.

6 Computation in Place

It is important to understand that quantum circuits differ fundamentally from their classical counterparts, such as printed circuit boards or integrated circuits. In a quantum circuit, the gates and wires are not tangible physical entities in the same way. Instead, only the qubits and their control mechanisms are physically implemented. The wires in the diagram represent the temporal continuity of qubit states, and the quantum gates denote specific transformations of these states, directed by external controls.

Exercise 5.12



- (a) What is the output state $|\psi\rangle$ after the above sequence of gates is applied to the initial state $|0\rangle$?

(b) Quantum circuits involving unitary gates are reversible. Draw the inverse circuit for the above sequence. Hint: the inverse of a circuit consists of the inverses of the original gates applied in reverse order.

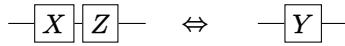
5.3.3 Gate Sequences

Equipped with our knowledge of quantum gates, we can delve into gate sequences and, crucially, their equivalences. Intriguingly, gate sequences can often be represented by a single, equivalent gate. This capability is vital for optimizing quantum circuits, especially when these circuits must be executed on real quantum hardware with resource constraints. Moreover, mastering the equivalences of gate sequences is foundational for analyzing complex quantum circuits.

In the subsequent discussions, we highlight some pivotal equivalences involving single-qubit gates. This list will be extended as we venture into multi-qubit gates in § 7.4.

1 Pauli Gate Relations

The diagram below demonstrates that the combination of a Z gate followed by an X gate is equivalent to a Y gate, up to a global phase: $ZX = iY$.



Further relations among the Pauli gates include: $XZ = -iY$, $XY = iZ$, $YX = -iZ$, $YZ = iX$, $ZY = -iX$. See Appendix D for more.

2 Involutory Relations

The diagram below shows that applying an H gate twice yields the identity operation (I), leaving the qubit state unchanged. Applying it thrice reduces to just applying it once, as a result of the involutory property of the Hadamard gate: $H^2 = I$.

$$\begin{array}{c} \xrightarrow{\hspace{1cm}} \boxed{H} \boxed{H} \xleftarrow{\hspace{1cm}} \\ \xrightarrow{\hspace{1cm}} \boxed{H} \boxed{H} \boxed{H} \xleftarrow{\hspace{1cm}} \end{array} \Leftrightarrow \begin{array}{c} \xrightarrow{\hspace{1cm}} \boxed{I} \xleftarrow{\hspace{1cm}} \\ \xrightarrow{\hspace{1cm}} \boxed{H} \xleftarrow{\hspace{1cm}} \end{array}$$

Other gates exhibit similar involutory relations: $X^2 = I$, $Y^2 = I$, and $Z^2 = I$.

3 Relations Among Hadamard and Pauli Gates

Given the relations $HXH = Z$ and $HX = ZH$, we derive the following gate sequence equivalences:

$$\begin{array}{c} \xrightarrow{\hspace{1cm}} \boxed{X} \boxed{H} \xleftarrow{\hspace{1cm}} \\ \xrightarrow{\hspace{1cm}} \boxed{H} \boxed{Z} \boxed{H} \xleftarrow{\hspace{1cm}} \\ \xrightarrow{\hspace{1cm}} \boxed{H} \boxed{X} \boxed{H} \xleftarrow{\hspace{1cm}} \end{array} \Leftrightarrow \begin{array}{c} \xrightarrow{\hspace{1cm}} \boxed{H} \boxed{Z} \xleftarrow{\hspace{1cm}} \\ \xrightarrow{\hspace{1cm}} \boxed{X} \xleftarrow{\hspace{1cm}} \\ \xrightarrow{\hspace{1cm}} \boxed{Z} \xleftarrow{\hspace{1cm}} \end{array}$$

4 Constructing the Hadamard Gate

In practice, the Hadamard gate is often realized using other basic gates. For instance, $H = iR_x(\pi)R_y(\frac{\pi}{2}) = iR_y(\frac{\pi}{2})R_z(\pi)$. This realization allows us to construct the Hadamard gate as illustrated below:

$$\xrightarrow{\hspace{1cm}} \boxed{R_y(\pi/2)} \boxed{R_x(\pi)} \xleftarrow{\hspace{1cm}} \Leftrightarrow \xrightarrow{\hspace{1cm}} \boxed{R_z(\pi)} \boxed{R_y(\pi/2)} \xleftarrow{\hspace{1cm}} \Leftrightarrow \xrightarrow{\hspace{1cm}} \boxed{H}$$

5 Constructing the Y Gate

In practice, the Y gate is often realized using other basic gates. For example, $Y = R_x(-\frac{\pi}{2})ZR_x(\frac{\pi}{2})$.

$$\xrightarrow{\hspace{1cm}} \boxed{R_x(\pi/2)} \boxed{Z} \boxed{R_x(-\pi/2)} \xleftarrow{\hspace{1cm}} \Leftrightarrow \xrightarrow{\hspace{1cm}} \boxed{Y}$$

Exercise 5.13 Draw circuit diagrams corresponding to the gate relations $HXH = Z$, $HYH = -Y$, and $HY = -YH$, $SXS^\dagger = Y$.

5.4 Quantum Random Number Generator

Quantum applications that use only single-qubit gates are relatively simple, as they do not involve entanglement or complex interactions between multiple qubits. Nonetheless, they can help us build an understanding of basic quantum concepts and operations. Next, we explore some examples, starting with QRNG.

A Quantum Random Number Generator (QRNG) is a device that generates truly random numbers by exploiting the inherent randomness and unpredictability of quantum mechanical processes. Unlike classical random number generators, which either rely on deterministic algorithms or on external sources of entropy, QRNGs harness the fundamental properties of quantum mechanics to produce genuine randomness that is independent of any underlying patterns or biases.

QRNG Using Photons

A common method to implement a QRNG is by using a single-photon source and a beam splitter, illustrated in Fig. 5.6(a). In this setup, individual photons are emitted and sent through the beam splitter. The beam splitter has a 50% chance of reflecting the photon and a 50% chance of transmitting it. Detectors are placed at both the reflected and transmitted paths, and the path taken by the photon is measured. At this juncture, we are measuring individual photons, positioning us firmly within the quantum regime, rather than the classical optics regime. Given that the photon's behavior at the beam splitter is inherently random due to quantum mechanics, the detector outcomes generate a truly random binary sequence.

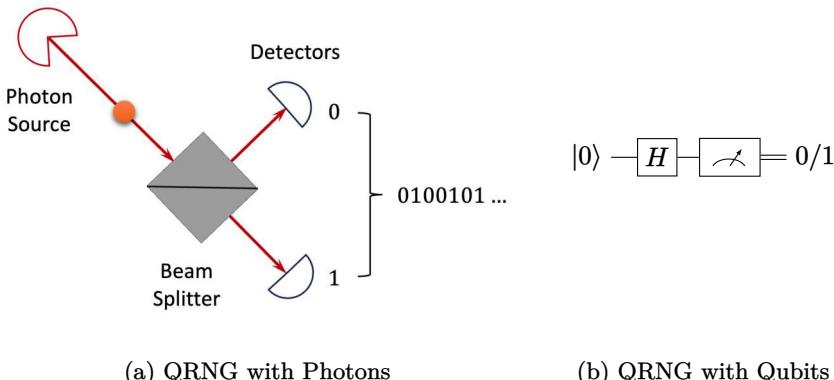


Figure 5.6: Quantum Random Number Generator (QRNG)

QRNG Using Qubits

Another approach to QRNGs is to exploit the quantum superposition and measurement process. A qubit is prepared in a superposition of its basis states, typically by applying a Hadamard gate to the initial state, as shown in Fig. 5.6(b). The qubit is then measured in the computational basis, collapsing the superposition to either 0 or 1 with equal probability. The measurement outcome serves as a random bit. The circuit is run repeated to produce a string of random bits of desired length.

Equivalence

The equivalence between these two methods can be understood by comparing the role of the beam splitter in the photon approach to that of the Hadamard gate in the superposition-measurement approach. The beam splitter places an incoming photon into a superposition of potential paths (reflected or transmitted), introducing quantum uncertainty. Similarly, the Hadamard gate places a qubit into a superposition of its basis states. When the photon or the qubit is observed or measured, this superposition results in inherent quantum randomness due to the unpredictability of the outcome. Both mechanisms, therefore, harness quantum superposition to generate random bits.

Exercise 5.14 * The QRNG devices discussed above generate binary sequences that correspond to random numbers with a uniform distribution. Discuss how you

would transform these uniformly distributed random numbers into numbers that follow a Gaussian distribution and a Poisson distribution. Consider the following in your discussion:

- (a) Describe any algorithms or mathematical transformations you would use to achieve these distributions from a uniform distribution.
- (b) Discuss any practical considerations or limitations that might arise when implementing these transformations, especially in the context of using a quantum device.
- (c) Discuss how the properties of quantum randomness impacts the quality or characteristics of the generated numbers compared to classical methods.

5.5 The BB84 Quantum Key Distribution (QKD) Protocol

Another application that primarily uses single-qubit gates is the BB84 Quantum Key Distribution (QKD) Protocol. As we will discuss next, this protocol serves as a beautiful showcase of the real-world application of quantum measurements and quantum gates.

5.5.1 Fundamentals of Cryptography

1 Cryptography

Cryptography comprises techniques for safeguarding communication and data against adversaries. Cryptographic systems fall into two main categories:

- Secret (or symmetric) key cryptography, where both parties utilize a shared key for encryption and decryption.
- Public (or asymmetric) key cryptography, which employs a dual-key system with public and private components. While data encrypted using the public key can only be decrypted using the private key, the significant overhead usually makes this method more suitable for exchanging secret keys.

2 Key Exchange Protocols

Key exchange protocols serve as intermediaries to facilitate the secure sharing of secret keys between parties. Classical key exchange protocols often employ asymmetric cryptographic techniques. Notable examples include:

- Diffie-Hellman key exchange is based on the difficulty of the discrete logarithm problem in a finite field. Both parties use their respective private and shared public parameters to generate an identical shared secret, which is computationally infeasible to deduce from the public parameters.
- RSA key exchange utilizes the RSA algorithm, an asymmetric cryptographic technique. The public key encrypts the shared secret, and only the private key can decrypt it. RSA relies on the computational difficulty of integer factorization.
- Elliptic Curve Diffie-Hellman (ECDH) key exchange operates similarly to the original Diffie-Hellman but employs elliptic curve cryptography, providing

equivalent security with shorter key lengths, thus being computationally more efficient.

- Lattice-based key exchange is a more recent approach, considered to be resistant to attacks from quantum computers, hence termed “quantum-safe.” These also usually employ asymmetric cryptographic principles.

It is worth mentioning that asymmetric cryptographic techniques, especially in protocols like RSA and Diffie-Hellman, are often used specifically for the key exchange process. Once the key has been securely exchanged, communication might then switch to a symmetric key cryptographic scheme, which is generally less computationally intensive.

3 Quantum Computing and Cryptography

The advent of quantum computing poses significant threats to traditional cryptographic protocols. Shor’s algorithm, for example, can efficiently factor large integers and compute discrete logarithms, effectively breaking RSA and Diffie-Hellman protocols. This vulnerability has led to an ongoing effort in the field of post-quantum cryptography, which aims to develop cryptographic protocols that are secure against quantum attacks. Some of the emerging options include lattice-based, code-based, and multivariate polynomial cryptography [14].

4 Quantum Key Distribution (QKD)

QKD leverages quantum mechanics to deliver a categorically secure public-key cryptosystem. QKD protocols can even detect potential eavesdroppers.

Notable QKD protocols include:

- BB84, proposed by Charles Bennett and Gilles Brassard in 1984, serves as one of the foundational QKD protocols.
- E91, which relies on entangled qubit pairs to establish a secure communication channel.
- B92, a less complex yet also less efficient variant of BB84.
- SARG04, an error-resilient protocol that draws inspiration from BB84.
- COW, which is compatible with existing telecommunications infrastructure and also takes inspiration from BB84.

5.5.2 Introduction to the BB84 QKD Protocol

The BB84 QKD protocol is renowned as the inaugural and a quintessential QKD protocol. It facilitates Alice and Bob in establishing a confidential shared key, suitable for secure communication via classical encryption methods. By exploiting quantum mechanics properties, including the no-cloning theorem, the BB84 protocol underscores the potential of quantum physics for bolstering communication security.

Within this protocol, Alice constructs a series of qubits (typically photons) with bases and states determined randomly. She transmits these qubits to Bob through a quantum channel, who then measures the received qubits using randomly opted bases. Subsequently, they utilize a public classical channel to compare their bases, retaining matches and dismissing mismatches. This shared set of measurement results becomes their secret key.

The protocol's security stems from the Heisenberg Uncertainty Principle (HUP, see § 1.6) and the no-cloning theorem (see § 5.1). Owing to the HUP, an eavesdropper cannot simultaneously and precisely measure both bases, given their non-orthogonal nature. The no-cloning theorem affirms the inability to perfectly replicate unknown quantum states, impeding an eavesdropper's attempt to intercept and relay qubits undetected. Consequently, any eavesdropping introduces errors, detectable during the error-checking phase, which alerts Alice and Bob to the need for restarting the transmission.

5.5.3 Procedure of the BB84 QKD Protocol

Typically, employing photon polarization, the BB84 protocol comprises five pivotal steps, elucidated in Fig. 5.7:

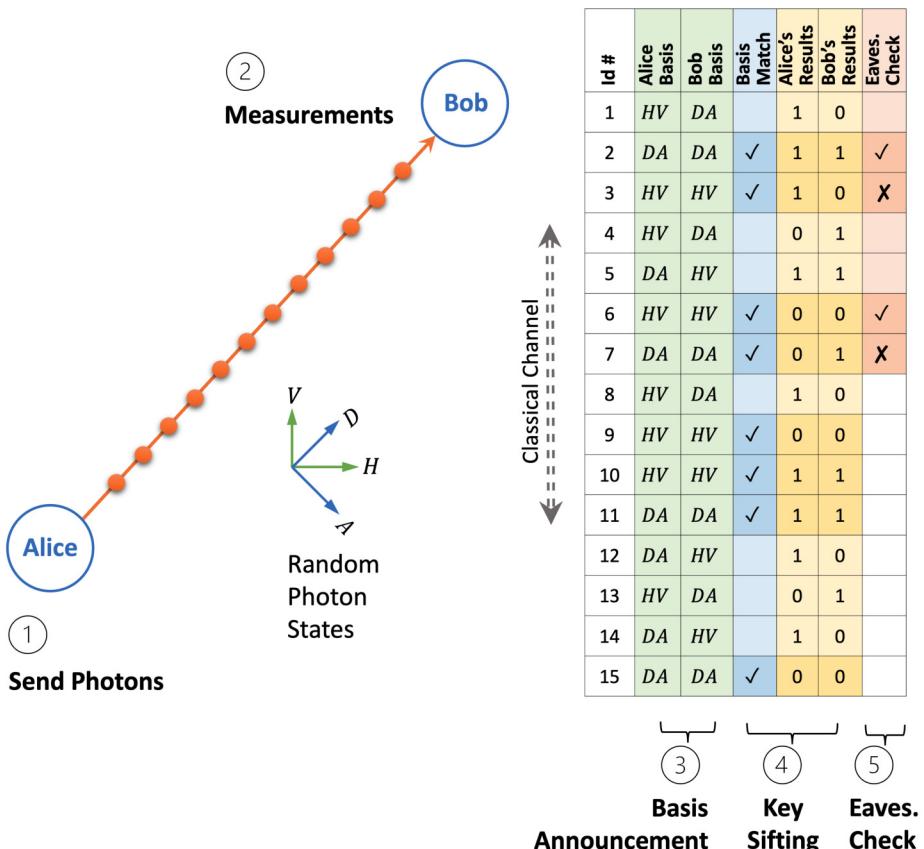


Figure 5.7: Fundamental BB84 QKD Protocol

1 Photon Generation and Distribution

Alice uses a QRNG to generate a random bit sequence. For each bit, she also randomly chooses either the rectilinear (HV) or diagonal (DA) basis. She then encodes the bits into photon polarizations based on the chosen bases:

- H polarization for bit 0 in the rectilinear basis

- V polarization for bit 1 in the rectilinear basis
- D polarization for bit 0 in the diagonal basis
- A polarization for bit 1 in the diagonal basis



The rectilinear basis, synonymous with the computational basis, aligns the H and V polarizations with $|0\rangle$ and $|1\rangle$ states, respectively. The diagonal basis mirrors the $\{|+\rangle, |-\rangle\}$ basis, associating the D (45°) polarization with $|+\rangle$ and the A (135°) polarization with $|-\rangle$.

Alice then dispatches the photons to Bob through a quantum medium, which could be a free space light beam or an optical fiber.

2 Measurement

Upon receiving each photon, Bob employs a randomly chosen basis (also either rectilinear or diagonal) to measure each photon's polarization. If Bob's basis aligns with Alice's, he discerns the intended bit. Otherwise, his measurement might diverge. This disparity is illustrated in Fig. 5.7.

3 Basis Announcement

Via a classical public channel, Bob notifies Alice of his measurement bases without disclosing the results.

4 Key Shifting

Comparing bases, Alice and Bob retain results from matching bases and eliminate the rest. They then split the remaining sifted bits for eavesdropping detection and as a shared secret key. Typically, these groups might span 128 and 2048 bits, respectively. However, Fig. 5.7 only illustrates this with 4 bits for each group.

5 Eavesdropping Check

Alice and Bob examine the first sifted key group to identify potential eavesdropping, as demonstrated in the column labeled "Eaves. Check" in Fig. 5.7. Any detected breach compels them to discard the key, initiating a new BB84 QKD session.

5.5.4 Detection of Eavesdropping

Eavesdropping refers to the act of intercepting or copying the quantum states of qubits (typically photons) during a quantum communication process. An eavesdropper, often referred to as Eve, may attempt to gain information about the transmitted key by measuring these qubits. However, the fundamental principles of quantum mechanics naturally limit her ability to do so without detection.

The Heisenberg's Uncertainty Principle (HUP) posits that one cannot measure both the rectilinear and diagonal polarization states of a qubit with arbitrary precision at the same time. This is due to the inherent non-orthogonality of the two bases. Consequently, when Eve measures the qubits using an incorrect basis, she invariably disturbs the states, introducing potential errors in the transmitted key.

Furthermore, the No-Cloning Theorem (see § 5.1.3) emphasizes that it is unfeasible to produce identical copies of an unknown quantum state. This theorem

considerably hinders Eve's chances of intercepting and replicating quantum states without alerting the communicating parties.

There are a few notable scenarios wherein Eve might try to eavesdrop:

1. Eve selects the incorrect basis, leading to an altered qubit:

- Should Bob opt for the correct basis, the introduced error by Eve becomes detectable.
- On the other hand, if Bob selects an incorrect basis, the result is inherently random, thus keeping Eve concealed.

2. Eve picks the correct basis, leaving the qubit unaffected:

- If Bob subsequently chooses the correct basis, Eve remains undetected while successfully acquiring one bit of the key.
- Conversely, if Bob selects the wrong basis, the outcome remains unpredictable, and Eve's presence stays hidden.

Given these conditions, the probability of detecting Eve when using n check bits can be expressed as:

$$P(\text{Detection of Eve with } n \text{ check bits}) = 1 - \left(\frac{3}{4}\right)^n. \quad (5.26)$$

For ensuring a security confidence level of 99.9%, Alice and Bob would need to implement error-checking by comparing a subset of 30 bits from the key to discern any potential eavesdropping activities.

Exercise 5.15 Describe the BB84 protocol for Quantum Key Distribution. Your description should include:

- (a) The process of key generation and distribution between Alice and Bob.
- (b) The role of the quantum and classical channels.
- (c) The method used to detect eavesdropping by a third party (Eve).
- (d) The procedure for key sifting and reconciliation.

5.5.5 BB84 QKD Implementation with Qubits

The BB84 protocol, which we have previously explained using photon polarization, can also be generalized to employ any qubit system in principle.

1 Key Steps

The key steps of the BB84 protocol using general qubits are outlined below, mirroring the same steps in the photon implementation.

1. Alice prepares qubits randomly in one of four states ($|0\rangle$, $|1\rangle$, $|+\rangle$, $|-\rangle$) and sends them to Bob. The key information is encoded in two bases:

- Computational basis: $\{|0\rangle, |1\rangle\}$, where $|0\rangle$ encodes bit 0 and $|1\rangle$ encodes bit 1.

- Hadamard basis: $\{|+\rangle, |-\rangle\}$, where $|+\rangle$ encodes bit 0 and $|-\rangle$ encodes bit 1.
2. Bob measures the received qubits using a randomly chosen basis.
 3. Alice and Bob reveal the bases they used for preparing and measuring the qubits through a classical communication channel.
 4. They keep the bits where both Alice and Bob used the same basis and discard the rest. The remaining bits form the shared secret key.
 5. They examine a segment of sifted key to identify potential eavesdropping.

2 Implementation with Quantum Gates

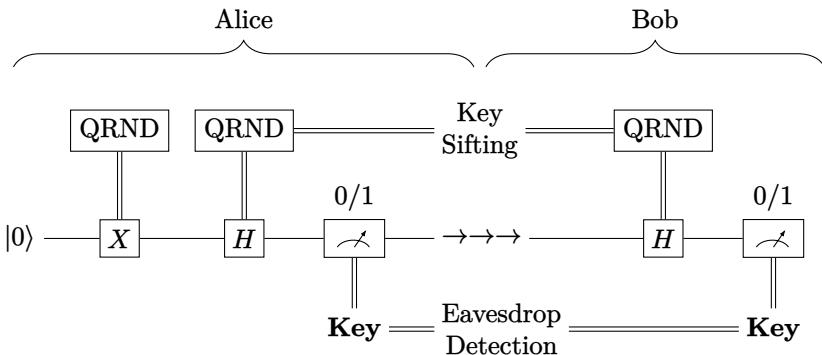


Figure 5.8: BB84 QKD Implementation with Qubits

In this generalized implementation, illustrated in Fig. 5.8, the BB84 protocol uses qubits in four states: $|0\rangle$, $|1\rangle$, $|+\rangle$, and $|-\rangle$. By convention, single lines represent qubits (or quantum channels), while double lines represent classical bits (or classical channels). Alice employs the following methods to prepare her qubit states:

- $|0\rangle$: The qubit is always initialized in the $|0\rangle$ state.
- $|1\rangle$: Apply an X gate after initialization.
- $|+\rangle$: Apply a Hadamard (H) gate after initialization.
- $|-\rangle$: Apply an X gate followed by an H gate after initialization.

Alice's X and H gates are controlled by two independent Quantum Random Number Generator (QRND) devices. A gate is switched into effect if the output bit from its corresponding QRND is 1; otherwise, it does not change the qubit state.

Similarly, when Bob receives a qubit, he selects a random basis using a QRND-controlled H gate. The circuit is meant to run repeatedly to produce a random binary string of desired length.

Exercise 5.16 Assume Eve is trying to eavesdrop on Alice and Bob during their BB84 protocol by performing an intercept-and-resend attack.

- (a) Explain Eve's strategy and how it might affect the key generation process.

- (b) What is the probability that Eve's interference will be detected by Alice and Bob?
- (c) Discuss how Alice and Bob can quantify and mitigate the information gained by Eve.
- (d) Express the eavesdrop-detection process using a quantum circuit.

5.5.6 *Further Exploration

The BB84 QKD protocol leverages the fundamental properties of quantum mechanics to establish a secure communication channel. While classical encryption techniques can be rendered ineffective by the advancement of computing capabilities (especially quantum computers), the security of the BB84 protocol is grounded in the immutable principles of quantum mechanics, making it a future-proof solution for secure communication.

However, practical implementations of the BB84 protocol do face challenges such as photon loss, noise, and equipment imperfections. Continuous research and development efforts are necessary to address these issues and ensure the successful deployment of QKD systems in real-world scenarios. For readers interested in exploring this subject further, the following resources are recommended:

1. General Introduction to Cryptography and Key Distribution Protocols [5]: This textbook provides a comprehensive introduction to modern cryptography, including key distribution protocols.
2. BB84: Original Paper by Charles Bennett and Gilles Brassard [21]: The original paper that introduced the BB84 protocol, laying the foundation for quantum key distribution.
3. B92: A Simpler Yet Less Efficient Variant of BB84 [22]: This paper introduces B92, a simplified and less efficient variant of the BB84 protocol.
4. SARG04: Error-Resilient Quantum Cryptography Inspired by BB84 [79]: A paper that introduces SARG04, an error-resilient quantum key distribution method inspired by BB84.
5. COW: Congruent with Existing Telecom Infrastructure and Inspired by BB84 [86]: A detailed exploration of the COW protocol, compatible with existing telecommunication infrastructure and influenced by BB84.
6. E91: Quantum Cryptography Using Entangled Qubit Pairs [39]: This paper presents the E91 protocol that utilizes entangled qubit pairs for key distribution. We will also discuss the E91 QKD protocol in § 10.6.
7. Quantum Key Distribution in Cryptography: A Survey [13]: This article explores the integration of quantum key distribution (QKD) in cryptographic infrastructures. Highlighting QKD's promise of information-theoretic security, the review underscores its application in renewing symmetric cipher keys and enabling secure key establishment in networks, while addressing inherent challenges and research avenues.
8. Post-Quantum Cryptography [14]: The vulnerabilities of traditional cryptographic protocols, such as RSA and Diffie-Hellman, have led to ongoing efforts in the field of post-quantum cryptography. This field aims to de-

velop cryptographic protocols that are secure against attacks leveraging Shor's algorithm.

5.6 The Quantum Coin Game

Prepare to explore an exciting game: the fair coin game, with Alice and Bob putting a quantum twist on it! This is not merely a game between these two players but an intriguing introduction to the world of quantum algorithms. You, the reader, are invited to follow along as we unravel the fascinating potentials of this quantum challenge.

5.6.1 The Game Rules

To ensure fairness and adaptability in the quantum version of the game, we will use the “Coin in Box” version of the game. Here are the game rules:

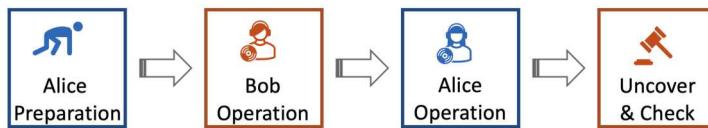


Figure 5.9: Fair Coin Gate - Setup and Rules

1. Alice Preparation: Alice places the coin in a state of her choice (heads or tails on top) in a box that covers the coin. (Note this is a special coin. Even by touching the coin, it is impossible to determine which side is up.)
2. Bob Operation: Bob reaches into the box and has the choice to flip the coin over or not. Alice cannot see what he is doing.
3. Alice Operation: Alice reaches into the box and performs an operation of her choice (flip or not flip). Bob cannot see what she is doing.
4. Uncover and Check: The coin is uncovered, and the result is read.
 - If it is Heads, Alice wins.
 - If it is Tails, Bob wins.

Is it a fair game? In the classical version, you bet.

Now let's consider the quantum version of the game:

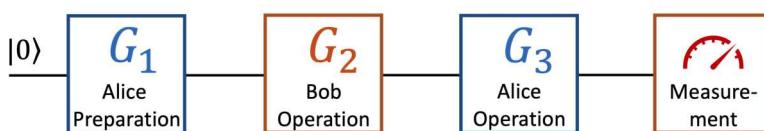


Figure 5.10: Quantum Coin Gate - Setup and Rules

The rules are the same as the classical version mentioned above. However, the following changes are made:

1. The coin is now represented by a qubit, with $|0\rangle$ representing Heads and $|1\rangle$ representing Tails.
2. A flip operation corresponds to the application of a NOT gate, which swaps the states $|0\rangle$ and $|1\rangle$.
3. A non-flip operation corresponds to the application of the Identity (I) gate.
4. In addition to the flip and non-flip operations, Alice and Bob are allowed to use other single-qubit gates (labeled G_1 , G_2 , and G_3 in Fig. 5.10) in their strategies.
5. In the Preparation step, the qubit is initially in state $|0\rangle$. However, Alice can apply a quantum gate to change it to another state.
6. The final Uncover and Check step is replaced by quantum measurement. If the measurement outcome is $|0\rangle$, Alice wins; if it is $|1\rangle$, Bob wins.

In this quantum version of the game, players can use the principles of quantum mechanics to devise new strategies and possibly gain an advantage over their opponent.

5.6.2 Alice Always Wins!

In the quantum case, Alice can win each time by using the Hadamard (H) gate. Here's the strategy:

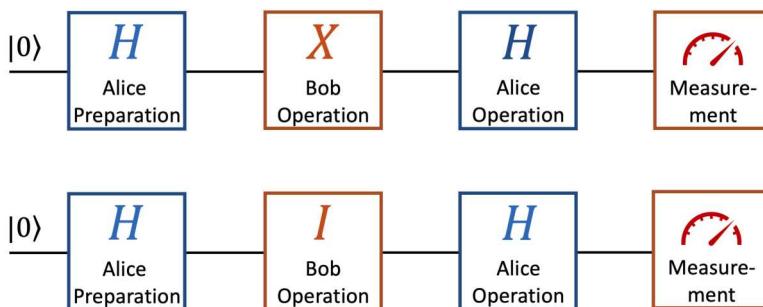


Figure 5.11: Strategy in Which Alice Always Wins

1. Alice applies the Hadamard gate, putting the qubit in an equal superposition state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, namely, $H|0\rangle = |+\rangle$.
2. Regardless of whether Bob applies the I gate or the X gate, the qubit remains in a superposition state. However, in either case, the qubit state is unchanged, because $I|+\rangle = |+\rangle$ and $X|+\rangle = |+\rangle$.
3. Alice applies the Hadamard gate again, converting $|+\rangle$ back to $|0\rangle$ since $H|+\rangle = |0\rangle$.
4. They measure the final state. However, since the final state will always be $|0\rangle$, Alice will win every time.

By introducing the Hadamard gate into the game strategically, Alice ensures that the final state of the qubit is always Heads ($|0\rangle$), allowing her to win the game every time!

5.6.3 Bob's Strategy Shift

Assuming Alice keeps her current strategy of using the Hadamard gate, Bob can adapt his strategy to change the game's outcome in his favor. Here's how Bob can use other gates to beat Alice:

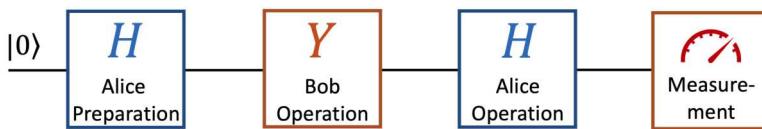


Figure 5.12: Strategy Shift: Bob Beats Alice

1. Alice applies the Hadamard gate, putting the qubit in an equal superposition state $|+\rangle$.
2. Bob applies the Pauli-Y gate to the qubit instead of X or I . After applying the Y gate, the qubit state becomes $-i\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, which is $|-\rangle$, since the global phase $(-i)$ can be ignored.
3. Alice applies the Hadamard gate again. The resulting qubit state is $|1\rangle$ because $H|-\rangle = |1\rangle$.
4. They measure the final state. Since the outcome is $|1\rangle$, Bob wins.

By using the Y gate strategically, Bob can change the outcome of the game and beat Alice!

5.6.4 Alice Overcomes Bob's New Gambit

If Bob continues using his new strategy with the Y gate, Alice can also revise her strategy to beat Bob again. How can that be done? She can compose a new gate G using a gate sequence.



Figure 5.13: Strategy Revised: Alice Wins Again

Exercise 5.17 Assuming Bob continues using his new strategy with the Y gate, Alice would revise her strategy to beat Bob again. What gate can she use to achieve this? Here is a hint: $XHYH|0\rangle = i|0\rangle$.

You can imagine that, since they have mastered quantum gates and sequences, Alice and Bob could revise their strategies continuously, playing happily thereafter.

5.7 *The No-Cloning Theorem: Proof Outline

The no-cloning theorem, as introduced in § 5.1.3, asserts that it is impossible to produce an exact replica of an arbitrary unknown quantum state using any physically realizable quantum operation. This section provides a proof of the theorem. Given that the proof involves concepts of multi-qubit gate operation, it is recommended that readers come back to this section after familiarizing themselves with these concepts in Chapter 7.

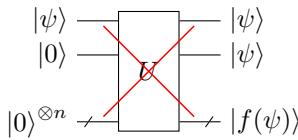


Figure 5.14: Illustration of the No-Cloning Theorem

Let's consider unitary operations first: a hypothetical quantum cloning machine represented by a unitary operator U . This operator acts on a quantum state $|\psi\rangle$, which we wish to clone, and an ancillary state $|0\rangle$, prepared for the cloning process. After applying U , our goal is to obtain two identical copies of $|\psi\rangle$.

Additionally, the cloning system may involve extra qubits. Assuming these qubits are initially in the $|0\rangle$ state without loss of generality, the cloning process can be described by:

$$U(|\psi\rangle \otimes |0\rangle \otimes |0\rangle^{\otimes n}) = |\psi\rangle \otimes |\psi\rangle \otimes |f(\psi)\rangle, \quad (5.27)$$

where $|f(\psi)\rangle$ represents the state of the extra qubits after cloning.

To demonstrate the impossibility of such cloning, we proceed by contradiction. Assume that the same operator U can clone the states $|0\rangle$, $|1\rangle$, and $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. This leads to the following equations:

$$U(|0\rangle \otimes |0\rangle \otimes |0\rangle^{\otimes n}) = |0\rangle \otimes |0\rangle \otimes |f(0)\rangle, \quad (5.28a)$$

$$U(|1\rangle \otimes |0\rangle \otimes |0\rangle^{\otimes n}) = |1\rangle \otimes |1\rangle \otimes |f(1)\rangle, \quad (5.28b)$$

$$U(|+\rangle \otimes |0\rangle \otimes |0\rangle^{\otimes n}) = |+\rangle \otimes |+\rangle \otimes |f(+)\rangle. \quad (5.28c)$$

By the linearity of U , adding the first two equations and multiplying by $\frac{1}{2}$ gives:

$$U(|+\rangle \otimes |0\rangle \otimes |0\rangle^{\otimes n}) = \frac{1}{2}(|00\rangle \otimes |f(0)\rangle + |11\rangle \otimes |f(1)\rangle). \quad (5.29)$$

However, the third equation implies:

$$U(|+\rangle \otimes |0\rangle \otimes |0\rangle^{\otimes n}) = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) \otimes |f(+)\rangle. \quad (5.30)$$

The right-hand sides of Eqs. 5.29 and 5.30 represent different quantum states, because measuring the first two qubits in the former yields two possible outcomes

($|00\rangle$ and $|11\rangle$), while the latter yields four different outcomes. This contradiction confirms that it is impossible to create an exact replica of an arbitrary unknown quantum state using only unitary quantum operations.

The proof of the No-Cloning Theorem, demonstrated for unitary transformations above, can be extended to include Completely Positive Trace-Preserving (CPTP) maps, which encompass measurements and other quantum operations. This extension is outlined in § 12.2.9. Thus, the No-Cloning Theorem is not limited to unitary operators alone; it establishes that it is impossible to create an identical copy of an arbitrary unknown quantum state using any physically realizable quantum operation.

5.8 Summary and Conclusions

Understanding Quantum Gates

Throughout this chapter, we navigated the integral concept of quantum gates, which are foundational to quantum computing. Just as Newtonian mechanics underpins satellite launches, core quantum mechanics principles play a pivotal role in shaping and managing quantum gates. We expanded on our prior knowledge of quantum mechanics and qubits, venturing into the world of quantum gates.

Our introduction began with single-qubit gates, giving us a grasp on their varied applications. This groundwork paves the way for deeper insights into multi-qubit gates, entanglement, quantum teleportation, and the E91 quantum key distribution in future discussions.

Distinguishing Quantum from Classical Gates

We emphasized the difference between quantum gates and their classical counterparts. Classical gates modify classical bits, while quantum gates act on qubits, allowing the creation of quantum circuits adept at processing and storing quantum information. These discussions shed light on the divergences and parallels between classical and quantum computing.

Our look into single-qubit gates, like the Pauli and Hadamard gates, showcased the extensive possibilities unlocked when these gates are paired together, enabling quantum circuits to address challenges daunting for classical systems.

Real-World Quantum Applications

We explored applications such as the Quantum Random Number Generator (QRNG) and the BB84 Quantum Key Distribution (QKD) Protocol. These applications highlighted the practical significance of quantum gates, enhancing our understanding of quantum measurements and their uses. The Quantum Coin Game served as a practical demonstration of the potential within quantum algorithms.

The Next Step: Multi-Qubit Systems

After solidifying our understanding of single-qubit gate principles and applications, we are ready to delve into multi-qubit systems. This forthcoming exploration will establish the foundation for multi-qubit gates and the enigmatic concept of quantum entanglement. While classical mechanics employs straightforward addition to describe systems, we will see how quantum mechanics uses the tensor product

to represent states of joint quantum systems, offering a more complex interaction framework.

Looking Ahead

The next chapter will spotlight multi-qubit systems, an essential preface to advanced topics such as multi-qubit gates and quantum entanglement. Our focus will be on the composition principle in quantum mechanics, offering clarity on how quantum systems are distinctively represented compared to the classical sum-based approach. We'll aim to understand the nuances of quantum particle interactions and their progression over time, preparing us for an in-depth study of multi-qubit gates in subsequent chapters.

Problem Set 5

- 5.1** Show that the Pauli Y operator acts as a flip gate for the states $|+\rangle$ and $|-\rangle$, ignoring the global phase (which is inconsequential to the physical state of the qubit). Specifically, demonstrate that $Y|+\rangle = -i|-\rangle$ and $Y|-\rangle = i|+\rangle$. Then, demonstrate this transformation on the Bloch sphere.

- 5.2** Validate the relation $ZX = iY$ using both matrix and bra-ket representations.

- 5.3** Validate the following relations. Here $S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$.

$$\begin{array}{lll} XXX = X, & XYX = -Y, & XZX = -Z, \\ YXY = -X, & YYY = Y, & YZY = -Z, \\ ZXZ = -X, & ZYZ = -Y, & ZZZ = Z, \\ SXS^\dagger = Y, & SYS^\dagger = -X, & Szs^\dagger = Z. \end{array}$$

- 5.4** Consider $H' = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

- (a) If $H' = UH$, where H is the Hadamard operator, find U .
(b) Find $H'|0\rangle$, $H'|1\rangle$, $H'|+\rangle$, and $H'|-\rangle$.

- 5.5** Show how the Hadamard gate, H , can be expressed as $iR_x(\pi)R_y(\frac{\pi}{2})$ and $iR_y(\frac{\pi}{2})R_z(\pi)$.

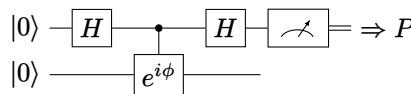
- 5.6** Show that, up to a global phase factor, $H = X\sqrt{Y}$ and $H = \sqrt{YZ}$.

- 5.7** Consider a qubit initially in the state defined by the polar angle α and the azimuthal angle β on the Bloch sphere, expressed as:

$$|\psi\rangle = \cos \frac{\alpha}{2} |0\rangle + e^{i\beta} \sin \frac{\alpha}{2} |1\rangle.$$

The rotation around the x -axis by an angle θ is represented by the matrix $R_x(\theta)$ (as defined in Eq. 5.17). Derive the state vector $|\psi'\rangle$ after the gate $R_x(\theta)$ acts on the state $|\psi\rangle$.

- 5.8** Verify that the Hadamard gate is equivalent to $R_u(\pi, \frac{\pi}{4}, 0)$, where R_u is defined in Eq. 5.25. Then, show that the Hadamard gate can be visualized as a 180° rotation around the $(x + z)$ -axis, which is the angular bisector between the x and z -axes.
- 5.9** Analyze the quantum circuit below and derive the measured probability P as a function of ϕ . Discuss how this circuit functions similarly to a Mach-Zehnder interferometer (§ 1.5.4.1).



- 5.10** Investigate the scenario where Alice and Bob are implementing the BB84 protocol, but their quantum channel is noisy, leading to a 1% error rate in the transmission of qubits.
- How does noise affect the key generation process?
 - Describe a method Alice and Bob could use to estimate the error rate in their key bits.
 - * Explain how Alice and Bob can still establish a secret key using error correction and privacy amplification techniques.

6. Multi-Qubit Systems

Contents

6.1	Systems of Two Qubits	139
6.1.1	Review of Single Qubit States	139
6.1.2	Two-Qubit States: Developing Intuition	140
6.1.3	General States vs. Product States	142
6.2	* Measurements of Two-Qubit Systems	143
6.2.1	* Local Measurements on Subsystems	143
6.2.2	* Joint Measurements	149
6.3	Multi-Qubit System States	151
6.3.1	The Composition Postulate of Quantum Mechanics	152
6.3.2	Basis States	152
6.3.3	Product States, Entangled States, and Correlated States	154
6.3.4	* Identical Particles	156
6.4	* Measurements of Multi-Qubit Systems	157
6.4.1	* Measurements on a Single Qubit	157
6.4.2	* Measurements on a Group of Qubits	158
6.4.3	* A Partial Measurement Paradigm for Quantum Algorithms	159
6.4.4	* Measurements in Alternate Bases	161
6.5	* Time Evolution of Multi-Qubit States	162
6.5.1	* Two-Qubit System	163
6.5.2	* Multi-Qubit System	163
6.5.3	* Example: The Ising Model	164
6.5.4	* Example: CNOT Gate Implementation	165
6.6	Summary and Conclusions	166
	Problem Set 6	167

In this chapter, we explore the exciting and complex realm of multi-qubit systems. This crucial understanding is a stepping stone towards more advanced topics in quantum computation, such as multi-qubit gates and quantum entanglement. The defining principle governing these composite quantum systems is the composition principle in quantum mechanics, a concept that markedly separates the quantum world from classical mechanics.

In classical Newtonian mechanics, individual subsystems and their interactions are described with straightforward addition. For instance, the combined effect of individual forces acting on a system can be depicted as the simple sum of these forces, as portrayed by the equation $\sum m\mathbf{a} = \sum \mathbf{F}$.

On the other hand, quantum mechanics approaches this differently. It utilizes state vectors and quantum operators to encapsulate the behavior and interactions of subsystems. According to the composition principle in quantum mechanics, the joint state of multiple quantum subsystems is captured by the tensor product of their individual states. This crucial distinction sets quantum mechanics apart from the additive principle that governs classical systems.

The tensor product principle allows us to understand the intricate dance of quantum particles and their time evolution. Moreover, it is an essential tool to describe entanglement, one of the most fascinating and puzzling phenomena of quantum physics.

In this chapter, we delve into the intriguing world of multiple qubits, setting the stage for a deeper investigation of multi-qubit gates in the subsequent chapter.

6.1 Systems of Two Qubits

Embarking on the journey of multi-qubit systems, we initially focus on the two-qubit system. The understanding gleaned from two-qubit systems will build our intuition and elucidate the more complex principles of multi-qubit systems, while concurrently demonstrating the underlying connection with single-qubit systems.

6.1.1 Review of Single Qubit States

The state of a single qubit can be represented as a two-dimensional complex vector:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \quad (6.1)$$

Mathematically, we express this as $|\psi\rangle \in \mathbb{C}^2$, with \mathbb{C}^2 symbolizing a complex Hilbert space of two dimensions. In the context of quantum computing, where we focus on finite dimensions, a Hilbert space is essentially a vector space that is equipped with an inner product.

The state vector $|\psi\rangle$ is normalized, which means:

$$\langle\psi|\psi\rangle = |\alpha|^2 + |\beta|^2 = \alpha\alpha^* + \beta\beta^* = 1. \quad (6.2)$$

When we measure the qubit in the basis $\{|0\rangle, |1\rangle\}$, we obtain $|0\rangle$ with probability $|\alpha|^2$ and $|1\rangle$ with probability $|\beta|^2$.

The vectors $|0\rangle$ and $|1\rangle$ form the standard (or computational) basis of the state space \mathbb{C}^2 . They are orthonormal, which means:

$$\langle i|j\rangle = \delta_{i,j}. \quad (6.3)$$

Here $i, j \in \{0, 1\}$, and $\delta_{i,j}$ is the Kronecker delta, which is defined as $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise.

6.1.2 Two-Qubit States: Developing Intuition

Now consider two independent qubits, A and B . Their state vectors are represented as

$$|\psi_A\rangle = \alpha|0_A\rangle + \beta|1_A\rangle, \quad |\psi_B\rangle = \gamma|0_B\rangle + \delta|1_B\rangle, \quad (6.4)$$

where $\alpha, \beta, \gamma, \delta$ are complex coefficients such that $|\alpha|^2 + |\beta|^2 = 1$ and $|\gamma|^2 + |\delta|^2 = 1$.

We can view the two qubits as a combined quantum system, which should be consistent with our view of them as two separate systems. Let's check this out.

According to the principles of quantum mechanics, the state vector of the combined two-qubit system, given that the two qubits are independent, is represented as the tensor product of the individual states:

$$|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\psi_B\rangle = (\alpha|0_A\rangle + \beta|1_A\rangle) \otimes (\gamma|0_B\rangle + \delta|1_B\rangle). \quad (6.5)$$

By expanding this equation, we find:

$$|\psi_{AB}\rangle = \alpha\gamma|0_A\rangle \otimes |0_B\rangle + \alpha\delta|0_A\rangle \otimes |1_B\rangle + \beta\gamma|1_A\rangle \otimes |0_B\rangle + \beta\delta|1_A\rangle \otimes |1_B\rangle. \quad (6.6)$$

1 Normalization Verification

To validate the quantum state $|\psi_{AB}\rangle$, let's confirm it is normalized:

$$\begin{aligned} \langle \psi_{AB} | \psi_{AB} \rangle &= |\alpha\gamma|^2 + |\alpha\delta|^2 + |\beta\gamma|^2 + |\beta\delta|^2 \\ &= (|\alpha|^2 + |\beta|^2)(|\gamma|^2 + |\delta|^2) \\ &= 1. \end{aligned} \quad (6.7)$$

2 Single-Qubit Measurement Verification

Suppose we measure qubit A and obtain $|0_A\rangle$. This process should select the terms with $|0_A\rangle$ in our state vector, i.e., $\alpha\gamma|0_A\rangle \otimes |0_B\rangle + \alpha\delta|0_A\rangle \otimes |1_B\rangle$. Interpreting the complex coefficients as probability amplitudes for $|0_A\rangle \otimes |0_B\rangle$ and $|0_A\rangle \otimes |1_B\rangle$, the associated probability is

$$\begin{aligned} P_{|0_A\rangle} &= |\alpha\gamma|^2 + |\alpha\delta|^2 \\ &= |\alpha|^2(|\gamma|^2 + |\delta|^2) \\ &= |\alpha|^2. \end{aligned} \quad (6.8)$$

This result matches our expectation when considering qubit A independently. The same method can be applied for measuring $|1_A\rangle$, $|0_B\rangle$, and $|1_B\rangle$. The probability interpretation remains consistent in all cases.

3 Two-Qubit Measurement Verification

Assume we measure qubit A and obtain $|0_A\rangle$, and simultaneously measure qubit B and obtain $|0_B\rangle$. The joint probability of these two measurement outcomes in independent systems would be $|\alpha|^2|\gamma|^2$. In a combined quantum system, these measurements select the term $\alpha\gamma|0_A\rangle \otimes |0_B\rangle$ from the state vector. The probability is given by the square of the coefficient, which also results in $|\alpha\gamma|^2$, thereby providing a consistent representation.

4 Orthonormal Basis Verification

To extend our state vector to represent generic two-qubit states, including interacting qubits, we require the terms $|0_A\rangle \otimes |0_B\rangle$, $|0_A\rangle \otimes |1_B\rangle$, etc., to form an orthonormal basis.

For this, we can express the basis states in matrix form:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (6.9)$$

By applying the tensor product rules for matrices, we get:

$$\begin{aligned} |0_A\rangle \otimes |0_B\rangle &= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, & |0_A\rangle \otimes |1_B\rangle &= \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \\ |1_A\rangle \otimes |0_B\rangle &= \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, & |1_A\rangle \otimes |1_B\rangle &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \end{aligned} \quad (6.10)$$

These equations clearly indicate that the set $\{|i_A\rangle \otimes |j_B\rangle\}$ ($i, j \in \{0, 1\}$) is orthonormal, meaning each pair of vectors is orthogonal, and each vector has unit length.

The two-qubit state space is four-dimensional. Since the above set contains four independent elements, it also spans the entire vector space and hence, it forms a complete orthonormal basis for this state space.

5 Tensor Product of Operators

In our discussion so far, we have successfully established that the state vector of two independent qubits can be represented as the tensor product of their individual state vectors. But how does this formalism extend to the operators that act on these state vectors?

Consider two operators: U_A acting on $|\psi_A\rangle$ and U_B acting on $|\psi_B\rangle$. For our combined two-qubit state, it is natural to anticipate the combined operator as the tensor product $U_A \otimes U_B$, resulting in the state $(U_A \otimes U_B)|\psi_{AB}\rangle$. This expectation is supported by the tensor product property:

$$(U_A \otimes U_B)(|\psi_A\rangle \otimes |\psi_B\rangle) = (U_A |\psi_A\rangle) \otimes (U_B |\psi_B\rangle). \quad (6.11)$$

This relationship confirms the anticipated behavior of tensor product operators on combined state vectors.

6 Shorthand Notation

For succinctness, we will use $|00\rangle$ instead of $|0_A\rangle \otimes |0_B\rangle$ from this point forward, where the first position represents qubit A , and the second position represents qubit B . Other combinations will follow the same convention. Therefore, our earlier state vector in Eq. 6.6 simplifies to:

$$|\psi_{AB}\rangle = \alpha\gamma|00\rangle + \alpha\delta|01\rangle + \beta\gamma|10\rangle + \beta\delta|11\rangle. \quad (6.12)$$

The four basis states are now represented as $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. Their orthonormal condition can be expressed as:

$$\langle ij|kl\rangle = \delta_{ik}\delta_{jl}, \quad (6.13)$$

where $i, j \in \{0, 1\}$, and δ_{ij} is the Kronecker delta function.

Exercise 6.1 Consider the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \frac{1}{\sqrt{5}}(|0\rangle - 2i|1\rangle)$$

- (a) Express $|\psi\rangle$ in terms of $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$.
- (b) Compute the probability of the first qubit being in state $|1\rangle$.
- (c) Compute the probability of the first qubit being in state $|0\rangle$ and the second qubit being in state $|1\rangle$.

6.1.3 General States vs. Product States

1 General Two-Qubit States

In our system of two qubits, the state space is four-dimensional with basis states $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. This allows us to construct a general two-qubit state as a linear combination of these basis states:

$$|\psi\rangle = \sum_{i,j \in \{0,1\}} c_{ij} |ij\rangle = c_{00} |00\rangle + c_{01} |01\rangle + c_{10} |10\rangle + c_{11} |11\rangle. \quad (6.14)$$

Here, c_{ij} are complex coefficients such that $\sum |c_{ij}|^2 = 1$. Therefore, a general state for a two-qubit system is a vector in the four-dimensional complex Hilbert space \mathbb{C}^4 .

2 Product States

The state $|\psi_{AB}\rangle \equiv |\psi_A\rangle \otimes |\psi_B\rangle$ referred to in Eq. 6.12 is a specific instance known as a product state.

A product state is defined as one where the state vector can be expressed as the tensor product of the state vectors of the individual qubits. It represents a system of independent qubits.

3 Non-Product States

Apart from product states, there exist states that cannot be expressed as a tensor product of individual qubit states. An example of such a state is the Bell state, also known as an EPR pair:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \quad (6.15)$$

This state cannot be written as a tensor product of individual qubit states.

Proof. If we assume that $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ could be expressed by Eq. 6.12, we find that $\alpha\gamma = \beta\delta = \frac{1}{\sqrt{2}}$ and $\alpha\delta = \beta\gamma = 0$. However, these two conditions are impossible to satisfy simultaneously. \square

Non-product states like these exhibit entanglement, a unique quantum phenomenon where the state of one qubit is interconnected with the state of the other, which we will study extensively later.

Exercise 6.2 Demonstrate that $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ is a non-product state.

Exercise 6.3 Prove that $\frac{1}{2}(|00\rangle - |10\rangle - |01\rangle + |11\rangle)$ is a product state and identify the two component states.

6.2 *Measurements of Two-Qubit Systems

We introduced a broad framework for quantum measurement in § 3.4. This section delves deeper by extending this framework to encompass the measurements of composite quantum systems, specifically those involving multiple qubits.

To provide a clear perspective, we will primarily focus on a composite quantum system comprising two subsystems, A and B . However, it's essential to note that the principles we discuss are readily extendable to more intricate systems consisting of multiple subsystems. These subsystems can be formed by qubits or other advanced quantum particles. When addressing the measurements of such composite systems, the tensor product structure of the state space becomes crucial, as do potential correlations between the subsystems. In the subsequent sections, we will elucidate the two predominant types of quantum measurements suited for these composite systems.

6.2.1 *Local Measurements on Subsystems

In quantum mechanics, a local measurement refers to a measurement performed on one subsystem of a composite system, leaving the other subsystem(s) physically undisturbed. The measurement operation applies only to the subsystem of interest, and the outcome pertains solely to the state of this subsystem.

For a composite system composed of subsystems A and B , the local measurement operators for a measurement on subsystem A or B can be expressed as $M_i \otimes I_B$ and $I_A \otimes M_i$ respectively, where M_i represents the measurement operators of the subsystem in focus and I_A and I_B are identity operators on subsystems A and B .

The sequence in which local measurements are made is not essential, as each measurement operates independently on its respective subsystem. Mathematically, $M_i \otimes I_B$ and $I_A \otimes M_i$ commute. The primary aim of local measurements is to uncover properties of the individual subsystem.

1 Entangled States vs. Product States

Entangled States

For entangled states, the state of the composite system cannot be fully described

by the states of its individual subsystems. When a measurement is made on one subsystem of an entangled pair, our description of the overall system, and hence of the other subsystem, must be updated. This can affect the potential outcomes of subsequent measurements on the other subsystem.

For instance, consider a two-qubit entangled state like the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. If a measurement on subsystem A yields the result $|0\rangle$, the entire system collapses into state $|00\rangle$, thus determining that any subsequent measurement on subsystem B would yield $|0\rangle$. Conversely, if the measurement on subsystem A results in $|1\rangle$, the entire system collapses into state $|11\rangle$, and a subsequent measurement on B would yield $|1\rangle$. This reflects the profound correlation that is characteristic of entangled states, underlining the fact that while no direct interaction occurs between the subsystems during a local measurement, the measurement can nevertheless affect our description and future measurement predictions of the other subsystem.

Product States

In contrast to entangled states, for product states such as $|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$, the system can be entirely described by the states of its individual subsystems, and the measurement outcome on one subsystem does not affect our understanding or predictions of the other subsystem. Furthermore, in this case, we can apply measurement operators directly to the individual subsystem states. For instance, $(M_i \otimes I_B) |\psi_{AB}\rangle$ simplifies to $(M_i |\psi_A\rangle) \otimes |\psi_B\rangle$.

If we perform a measurement on subsystem A and find it in some state $|\psi_A\rangle$, this does not impact the state or potential measurement outcomes for subsystem B , which remains in its original state $|\psi_B\rangle$. Consequently, for product states, the measurements on individual subsystems are truly independent, and each subsystem can be analyzed separately without the need to consider their correlations. This decoupling of subsystems is a hallmark of product states. It illustrates that even though the universe is composed of many quantum systems, we can treat each one independently if it is isolated from the rest.

2 Local Measurements in the Computational Basis

To gain a deeper understanding and operational proficiency on measurements on composite quantum systems, it is necessary to study the related mathematics. Let's start with local measurements in the computational (or standard) basis.

When measuring in the computational basis, the single-qubit measurement operators corresponding to the outcomes $|0\rangle$ and $|1\rangle$ are:

$$M_i = |i\rangle\langle i|, \quad (6.16)$$

where $i \in \{0, 1\}$. Note that these are projection operators with the property $M_i^2 = M_i$.

The two-qubit measurement operators for measuring on qubit A and qubit B are:

$$M_{iA} = |i\rangle\langle i| \otimes I, \quad M_{iB} = I \otimes |i\rangle\langle i|. \quad (6.17)$$

When performing local measurements on the general two-qubit state given by

$$|\psi\rangle = \sum_{i,j \in \{0,1\}} c_{ij} |ij\rangle, \quad (6.14)$$

we first apply the two-qubit measurement operators on $|\psi\rangle$:

$$M_{0A} |\psi\rangle = (|0\rangle\langle 0| \otimes I) \sum_{i,j \in \{0,1\}} c_{ij} |ij\rangle = c_{00} |00\rangle + c_{01} |01\rangle, \quad (6.18a)$$

$$M_{1A} |\psi\rangle = (|1\rangle\langle 1| \otimes I) \dots = c_{10} |10\rangle + c_{11} |11\rangle, \quad (6.18b)$$

$$M_{0B} |\psi\rangle = (I \otimes |0\rangle\langle 0|) \dots = c_{00} |00\rangle + c_{10} |10\rangle, \quad (6.18c)$$

$$M_{1B} |\psi\rangle = (I \otimes |1\rangle\langle 1|) \dots = c_{01} |01\rangle + c_{11} |11\rangle. \quad (6.18d)$$

Then, we can calculate the associated probabilities according to Eq. 3.38:

$$P_{0A} = \|M_{0A} |\psi\rangle\|^2 = \|c_{00} |00\rangle + c_{01} |01\rangle\|^2 = |c_{00}|^2 + |c_{01}|^2, \quad (6.19a)$$

$$P_{1A} = \dots = |c_{10}|^2 + |c_{11}|^2, \quad (6.19b)$$

$$P_{0B} = \dots = |c_{00}|^2 + |c_{10}|^2, \quad (6.19c)$$

$$P_{1B} = \dots = |c_{01}|^2 + |c_{11}|^2. \quad (6.19d)$$

Next, we will focus on measuring qubit A and obtaining $|0\rangle$. Other scenarios are similar. The post-measurement state is calculated according to Eq. 3.41:

$$|\psi_{0A}\rangle = \frac{M_{0A} |\psi\rangle}{\sqrt{P_{0A}}} = \frac{c_{00} |00\rangle + c_{01} |01\rangle}{\sqrt{|c_{00}|^2 + |c_{01}|^2}} = \frac{|0\rangle \otimes (c_{00} |0\rangle + c_{01} |1\rangle)}{\sqrt{|c_{00}|^2 + |c_{01}|^2}}. \quad (6.20)$$

This clearly shows the system collapses to a product state upon measurement, with qubit A in $|0\rangle$ and qubit B in $|\psi_{B|0A}\rangle = \hat{N}(c_{00} |0\rangle + c_{01} |1\rangle)$, where \hat{N} represents normalization, in this case, dividing by $\sqrt{|c_{00}|^2 + |c_{01}|^2}$.

The state of qubit B can also be calculated through the partial product:

$$|\psi_{B|0A}\rangle = \hat{N} \langle 0_A | \psi \rangle, \quad (6.21)$$

where the index A in $\langle 0_A |$ indicates that it interacts only with the state of qubit A in $|\psi\rangle$ through inner product, while leaving the state of qubit B unchanged. (In matrix operation, $\langle 0_A | \psi \rangle$ is equivalent to $(\langle 0 | \otimes I) |\psi\rangle$.)

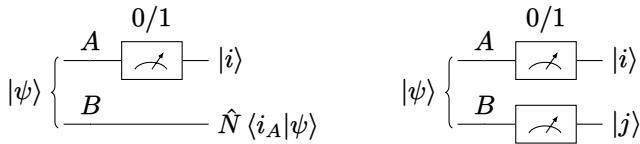
At this point, if we measure qubit B and get outcome $|1\rangle$, the system will collapse to $|01\rangle$. The probability for this outcome (A getting $|0\rangle$ and B getting $|1\rangle$) is

$$P_{1B,0A} = P_{0A} \left| c_{01} / \sqrt{|c_{00}|^2 + |c_{01}|^2} \right|^2 = |c_{01}|^2. \quad (6.22)$$

As you can see, the joint probability for the outcome $|01\rangle$ depends only on c_{01} as expected, and it does not depend on the order of the measurements on qubits A and B .

The results so far is illustrated in Fig. 6.1.

If we assign the value $\lambda_0 = 1$ to $|0\rangle$ and $\lambda_1 = -1$ to $|1\rangle$, as is the case when measuring spin, we can calculate the expected value (statistical average) of λ using Eq. 3.40:



Notations: $i, j \in \{0, 1\}$, and \hat{N} denotes for normalization.

Figure 6.1: Local Measurement Examples

$$\langle \lambda_A \rangle = \sum_i P_{iA} \lambda_i = |c_{00}|^2 + |c_{01}|^2 - |c_{10}|^2 - |c_{11}|^2, \quad (6.23a)$$

$$\langle \lambda_B \rangle = \sum_i P_{iB} \lambda_i = |c_{00}|^2 + |c_{10}|^2 - |c_{01}|^2 - |c_{11}|^2. \quad (6.23b)$$

In this case, the measurement corresponds to the Pauli Z operator observable, allowing us to calculate $\langle \lambda \rangle$ as:

$$\langle \lambda_A \rangle = \langle \psi | Z \otimes I | \psi \rangle, \quad (6.24a)$$

$$\langle \lambda_B \rangle = \langle \psi | I \otimes Z | \psi \rangle. \quad (6.24b)$$

Exercise 6.4 Consider the state

$$|\psi\rangle = c(|00\rangle - 2i|01\rangle + 2i|10\rangle + 4|11\rangle).$$

- (a) Determine the coefficient c to normalize the state $|\psi\rangle$.
- (b) Compute the probability of the first qubit being in state $|1\rangle$.
- (c) Compute the expected value for the observable $\frac{X+Z}{2}$.

3 Local Measurements in Alternate Bases

After familiarizing ourselves with the mechanics of measurements in the computational basis, we will now extend our discussion to local measurements in other bases. We will focus on one of the most common alternative bases: the Hadamard basis, with basis states $|+\rangle$ and $|-\rangle$, defined as:

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad (6.25a)$$

$$|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \quad (6.25b)$$

Consider the following observable:

$$M = |+\rangle\langle +| - |-\rangle\langle -|. \quad (6.26)$$

We can verify that M is a Hermitian operator, thereby qualifying it as an observable. When we use M to measure a qubit, the outcome will be either $|+\rangle$ or $|-\rangle$, the eigenstates of M . This observable associates the eigenvalue $+1$ with the outcome $|+\rangle$, and the eigenvalue -1 with the outcome $|-\rangle$.

Let's now use M to measure one of the Bell states, which are entangled states:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \quad (6.27)$$

To facilitate this, we will use the following properties:

$$\langle +|0\rangle = \langle +|1\rangle = \langle -|0\rangle = \frac{1}{\sqrt{2}}, \quad \langle -|1\rangle = -\frac{1}{\sqrt{2}}. \quad (6.28)$$

We apply M to the first qubit. In this process, we first derive the following intermediate quantities (un-normalized states):

$$|\psi_{+A}\rangle \equiv (|+\rangle\langle +| \otimes I)|\psi\rangle \quad (6.29a)$$

$$= \frac{1}{\sqrt{2}}(|+\rangle\langle +| |0\rangle I |0\rangle + |+\rangle\langle +| |1\rangle I |1\rangle) \quad (6.29b)$$

$$= \frac{1}{2} |+\rangle \otimes (|0\rangle + |1\rangle) \quad (6.29c)$$

$$= \frac{1}{\sqrt{2}} |++\rangle. \quad (6.29d)$$

Similarly,

$$|\psi_{-A}\rangle \equiv (|-\rangle\langle -| \otimes I)|\psi\rangle = \frac{1}{\sqrt{2}} |--\rangle. \quad (6.29e)$$

The probability of obtaining the outcome $|+\rangle$ is $\| |\psi_{+A}\rangle \| ^2 = \frac{1}{2}$. Upon measuring and obtaining the outcome $|+\rangle$, the state of the first qubit collapses to $|+\rangle$. Simultaneously, the state of the second qubit also collapses to $|+\rangle$, resulting in the two-qubit system being in the product state $|++\rangle$. Likewise for the outcome $|-\rangle$.

The expected value of M when measured on the first qubit is:

$$\langle \psi | M \otimes I | \psi \rangle = \langle \psi | (|\psi_{+A}\rangle - |\psi_{-A}\rangle) \rangle = 0. \quad (6.30)$$

Exercise 6.5 Consider $M = a |+\rangle\langle +| + b |-\rangle\langle -|$, and $|\psi\rangle = \cos \alpha |00\rangle + \sin \alpha |11\rangle$, where $a, b, \alpha \in \mathbb{R}$.

- (a) Show that M is an observable.
- (b) Describe the measurement outcomes, associated probabilities, and post-measurement states of the second qubit when the first qubit is measured with M .
- (c) Compute the expected value of M : $\langle \psi | M | \psi \rangle$.

In quantum computing, measurements are typically conducted in the computational basis due to its native support on quantum hardware. To measure $|\psi\rangle$ in the $|+\rangle, |-\rangle$ basis, we first need to convert it to the computational basis. This can be achieved by applying Hadamard gates:

$$H |+\rangle = |0\rangle, \quad H |-\rangle = |1\rangle, \quad (6.31a)$$

$$H |0\rangle = |+\rangle, \quad H |1\rangle = |-\rangle. \quad (6.31b)$$

After this transformation, we perform local measurements on the two qubits. For instance, if the measurement readouts are both 0, we deduce that the original state in the $|+\rangle, |-\rangle$ basis was $|++\rangle$. This logic can be similarly applied to other sets of readouts. The process is illustrated in Fig. 6.2.

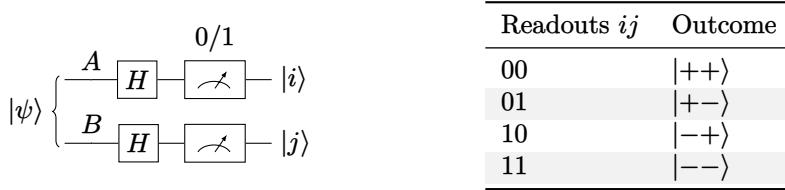


Figure 6.2: Local Measurement in $\{|+\rangle, |-\rangle\}$ Basis

4 * An Advanced Example



This example features advanced mathematical concepts and techniques. While first-time learners may opt to skip it initially, revisiting this content is recommended for a deeper understanding of Bell inequalities discussed in subsequent sections.

Consider qubit A measured along a direction defined by the polar angle α and the azimuthal angle θ on the Bloch sphere. Effectively, it is measured using the basis $\{|a\rangle, |a_\perp\rangle\}$:

$$|a\rangle = \cos \frac{\alpha}{2} |0\rangle + \sin \frac{\alpha}{2} |1\rangle, \quad (6.32a)$$

$$|a_\perp\rangle = -\sin \frac{\alpha}{2} |0\rangle + \cos \frac{\alpha}{2} |1\rangle. \quad (6.32b)$$

The associated observable is:

$$M_a = |a\rangle\langle a| - |a_\perp\rangle\langle a_\perp|. \quad (6.33)$$

Similarly, qubit B is measured along a different direction defined by the polar angle β on the Bloch sphere. The measurement basis $\{|b\rangle, |b_\perp\rangle\}$ is:

$$|b\rangle = \cos \frac{\beta}{2} |0\rangle + \sin \frac{\beta}{2} |1\rangle, \quad (6.34a)$$

$$|b_\perp\rangle = -\sin \frac{\beta}{2} |0\rangle + \cos \frac{\beta}{2} |1\rangle. \quad (6.34b)$$

The associated observable is:

$$M_b = |b\rangle\langle b| - |b_\perp\rangle\langle b_\perp|. \quad (6.35)$$

Let's consider the combined observable which represents the product (or correlation) of two local measurements:

$$M_{ab} = M_a \otimes M_b. \quad (6.36)$$

To measure the state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, we start with the expected values of M_a across the one-qubit basis states:

$$\langle 0 | M_a | 0 \rangle = \cos \alpha, \quad (6.37a)$$

$$\langle 1 | M_a | 1 \rangle = -\cos \alpha, \quad (6.37b)$$

$$\langle 1 | M_a | 0 \rangle = \sin \alpha, \quad (6.37c)$$

$$\langle 0 | M_a | 1 \rangle = \sin \alpha. \quad (6.37d)$$

Now calculate the expected values of M_{ab} across the two-qubit basis states:

$$\langle 11 | (M_a \otimes M_b) | 11 \rangle = \langle 1 | M_a | 1 \rangle \langle 1 | M_b | 1 \rangle = \cos \alpha \cos \beta, \quad (6.38a)$$

$$\langle 00 | (M_a \otimes M_b) | 00 \rangle = \langle 0 | M_a | 0 \rangle \langle 0 | M_b | 0 \rangle = \cos \alpha \cos \beta, \quad (6.38b)$$

$$\langle 11 | (M_a \otimes M_b) | 00 \rangle = \langle 1 | M_a | 0 \rangle \langle 1 | M_b | 0 \rangle = \sin \alpha \sin \beta, \quad (6.38c)$$

$$\langle 00 | (M_a \otimes M_b) | 11 \rangle = \langle 0 | M_a | 1 \rangle \langle 0 | M_b | 1 \rangle = \sin \alpha \sin \beta. \quad (6.38d)$$

Finally, we obtain:

$$\langle \psi | M_{ab} | \psi \rangle = \cos \alpha \cos \beta + \sin \alpha \sin \beta = \cos(\alpha - \beta). \quad (6.39)$$

6.2.2 *Joint Measurements

In a joint measurement, the measurement operator simultaneously acts on all subsystems under consideration. For instance, in a two-qubit system, we may measure a property that involves both qubits, and the measurement operator is a 4×4 matrix. The probabilities and post-measurement states are calculated by applying the joint measurement operator to the entire composite system state.

1 Parity Measurement

Consider two qubits in a system. They can be in the states $|00\rangle$, $|01\rangle$, $|10\rangle$, or $|11\rangle$. Our task is to measure the system and distinguish $|00\rangle$ and $|11\rangle$ as a group, and $|01\rangle$ and $|10\rangle$ as another, while leaving the state of the system unchanged.

This objective can be achieved using a parity measurement, which determines whether the total count of $|1\rangle$ states is even or odd. To be specific, $|00\rangle$ and $|11\rangle$ have even parity since there are either 0 or 2 $|1\rangle$ states. Conversely, $|01\rangle$ and $|10\rangle$ have odd parity due to the presence of one $|1\rangle$ state.

The parity measurement operator, when applied to the combined state of two qubits, is:

$$P_Z = Z \otimes Z = |00\rangle\langle 00| + |11\rangle\langle 11| - |01\rangle\langle 01| - |10\rangle\langle 10|. \quad (6.40)$$

The states $|00\rangle$ and $|11\rangle$ are eigenvectors of P_Z with an eigenvalue of 1, whereas $|01\rangle$ and $|10\rangle$ are eigenvectors with an eigenvalue of -1 . Thus, a measurement reading of 1 indicates even parity and -1 indicates odd parity. Since these states are eigenvectors of P_Z , the measurement leaves them unchanged.

Building on this, the mechanism can also be extended to superposition states. Consider the two superposition states:

$$|\psi_{\text{even}}\rangle = \alpha |00\rangle + \beta |11\rangle, \quad (6.41a)$$

$$|\psi_{\text{odd}}\rangle = \gamma |01\rangle + \delta |10\rangle. \quad (6.41b)$$

These states are also eigenvectors of P_Z :

$$P_Z |\psi_{\text{even}}\rangle = |\psi_{\text{even}}\rangle, \quad (6.42\text{a})$$

$$P_Z |\psi_{\text{odd}}\rangle = -|\psi_{\text{odd}}\rangle. \quad (6.42\text{b})$$

When measuring a general state of the form $|\psi\rangle = \sum_{i,j \in \{0,1\}} c_{ij} |ij\rangle$ using P_Z , the state collapses to either $|\psi_{\text{even}}\rangle$ or $|\psi_{\text{odd}}\rangle$, corresponding to measurement readings of 1 or -1 , respectively. However, if the system is already in $|\psi_{\text{even}}\rangle$ or $|\psi_{\text{odd}}\rangle$, the P_Z measurement leaves the state undisturbed, highlighting the essence of parity measurement.

i Parity measurements can discern groups of qubit states because the measurement operator has degenerate eigenvalues, leading to corresponding subspaces in its eigenspace. Specifically, the degenerate eigenvalues of 1 and -1 correspond to two different subspaces: one spanned by the even parity states $|00\rangle$ and $|11\rangle$, and the other by the odd parity states $|01\rangle$ and $|10\rangle$. These subspaces are used to classify the qubits into groups without changing their states.

Analogous to P_Z , there is a phase-parity measurement, which measures in the $\{|+\rangle, |-\rangle\}$ basis and counts the number of $|+\rangle$ states. Its operator is represented as:

$$P_X = X \otimes X = |++\rangle\langle++| + |--\rangle\langle--| - |+-\rangle\langle+-| - |-+\rangle\langle-+|. \quad (6.43)$$

Parity measurements are crucial in numerous quantum algorithms and protocols. For instance, they are foundational in quantum error correction for detecting both bit and phase flip errors. Additionally, they are employed in certain quantum teleportation protocols, wherein the parity of two qubits informs the requisite correction operation for the teleported qubit.

To execute a parity measurement on an actual quantum device, one generally needs the capability to conduct controlled operations (such as a CNOT gate) to generate entanglement between the qubits. Subsequent individual qubit measurements then yield the parity of the initial state. This will be explored further in the context of error correction in § 12.4.

2 Bell Measurement

A Bell measurement, also known as a Bell state measurement or entanglement measurement, is a type of quantum measurement that is particularly important for quantum information processing tasks such as quantum teleportation and superdense coding. In essence, a Bell measurement is a joint measurement on a two-qubit system that distinguishes between the four mutually orthogonal entangled states, called Bell states. More specifically, a Bell measurement results in one of four possible outcomes and transforms the state of the two qubits into one of the four Bell states corresponding to the measurement outcome.

While a Bell measurement can be described theoretically, performing a Bell measurement on a real quantum computer is non-trivial. It requires a sequence of quantum gates that effectively transform the Bell basis into the computational basis. This typically involves a CNOT gate, with the first qubit as the control and the second as the target, followed by a Hadamard gate applied to the first qubit. Finally,

individual measurements are performed on each qubit. The measurement results then allow us to determine which of the four Bell states the system was in. We will explore this further in Chapter 8 when we investigate entanglement and Bell states in detail.

Exercise 6.6 This is a key exercise that tests your basic understanding of quantum measurements. Ensure you can complete it independently.

Consider a system of three qubits (A , B , and C), in the state

$$|\psi\rangle = \sum_{i,j,k \in \{0,1\}} c_{ijk} |ijk\rangle, \quad (6.44)$$

where $|ijk\rangle$ represents the joint computational basis states of the qubits A , B , and C , and c_{ijk} are complex coefficients.

- (a) Calculate the probability of measuring the third qubit (C) in the state $|0\rangle$.
- (b) If qubit C is measured and collapses to $|0\rangle$, determine the resulting joint state of qubits A and B . Also, describe the state of the three-qubit system.
- (c) Assuming a Bell measurement is performed on qubits A and B and the outcome is $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, find the probability of this outcome and describe the post-measurement state of the three-qubit system.

6.3 Multi-Qubit System States

Having introduced the fundamental concepts of composite quantum systems with two-qubit examples, we will now expand our discussion to encompass n -qubit systems, also referred to many-body systems in quantum mechanics. The dimension of the state space for such a system is $N = 2^n$, and correspondingly, an operator in this space would be an $N \times N$ matrix. This is a major factor contributing to the potential exponential computational capacity of quantum computers.

How large can n and N be in practical quantum computing? When $n = 300$, we find that $N = 2^{300}$, which is approximately 10^{90} . This figure surpasses the estimated number of atoms in our observable universe. As of 2024, quantum computers with up to $n = 1000$ qubits already exist. Theoretically, these computers possess immense computational capacity. However, in practice, they are plagued by errors and instability due to qubit decoherence. Facilitating error correction necessitates additional qubits. It is estimated that to have a quantum computer with 1000 error-free logical qubits, we would need approximately a million ($n = 10^6$) physical qubits.

The Role of Entanglement in Quantum Computing

In an n -qubit system, if all qubits are independent, their collective product state can be described using just $2n$ complex numbers. However, when the qubits are entangled, describing their state requires up to 2^n complex numbers. This significant increase in representational complexity is a key to why quantum computers can potentially achieve substantial speedups over classical computers.

Thus, entanglement is not just a feature of quantum systems; it is a crucial prerequisite for the enhanced computational power of quantum computing.

6.3.1 The Composition Postulate of Quantum Mechanics

Quantum mechanics leverages state vectors and quantum operators to delineate the behavior of quantum systems and their interactions. The Composition Postulate in quantum mechanics stipulates that the collective state of two or more quantum systems is described by the tensor product of their individual states. More succinctly, the joint state space of two or more quantum systems is the tensor product of their individual state spaces.

Postulate 5: Composite Systems

When two or more quantum systems are combined, their joint state space is described by the tensor product of the state spaces of the constituent systems.

The tensor product provides a mechanism to combine two Hilbert spaces, \mathcal{H}_A and \mathcal{H}_B , into a larger Hilbert space, denoted as \mathcal{H}_{AB} . As per the Composition Postulate, the joint state space of the combined system AB is described by the tensor product of the individual state spaces:

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B. \quad (6.45)$$

This implies that the basis states of \mathcal{H}_{AB} are derived from the tensor product of the basis states of the individual spaces. Let's say the basis states for system A are $|a_i\rangle$ and for system B are $|b_j\rangle$. Consequently, the basis states for the combined system AB are:

$$|a_i b_j\rangle = |a_i\rangle \otimes |b_j\rangle. \quad (6.46)$$

If the dimension of \mathcal{H}_A is m , and that of \mathcal{H}_B is n , the resulting dimension of \mathcal{H}_{AB} is mn .

6.3.2 Basis States

For an n -qubit system, the state space is described by the tensor product of the individual qubit state spaces. A single qubit is described by a two-dimensional complex Hilbert space. Therefore, an n -qubit system will be described by an N -dimensional complex Hilbert space, where $N = 2^n$.

The basis for an n -qubit system is a set of orthonormal states that span the Hilbert space of the composite quantum system. A standard choice for the basis of an n -qubit system is the computational basis (also known as the standard basis). The computational basis is formed by taking the tensor product of the individual qubit basis states, as follows:

For a single qubit, the basis states are $|0\rangle$ and $|1\rangle$. For n -qubits, take the tensor product of the individual qubit basis states to create a set of N orthonormal basis states.

1 Binary Representation

In the binary representation, each basis state corresponds to a binary string of length n , with each digit representing a qubit in the state $|0\rangle$ or $|1\rangle$. For example, for a two-qubit system, the basis states are $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. The computational basis states for an n -qubit system are represented by:

$$|0 \cdots 00\rangle, |0 \cdots 01\rangle, |0 \cdots 10\rangle, \dots, |1 \cdots 11\rangle, \quad (6.47)$$

where each binary string corresponds to a unique basis state.

These binary strings can be thought of as names, or tags, of the basis states. The vector of each basis state is a column vector of dimension N , with a ‘1’ at the position indexed by the corresponding binary number. For example,

$$\begin{aligned} |0 \cdots 00\rangle &= [1 \ 0 \ \cdots \ 0 \ 0]^T, \\ |1 \cdots 11\rangle &= [0 \ 0 \ \cdots \ 0 \ 1]^T. \end{aligned}$$

These basis states are orthonormal, because each basis state has exactly one ‘1’ in it, each at a different position.

A general n -qubit state can be written as:

$$|\psi\rangle = \sum_{x_1, x_2, \dots, x_n \in \{0,1\}} c_{x_1 x_2 \dots x_n} |x_1 x_2 \dots x_n\rangle, \quad (6.48)$$

or in this compact form:

$$|\psi\rangle = \sum_{x \in \{0,1\}^n} c_x |x\rangle. \quad (6.49)$$

This binary representation is particularly useful when working with the tensor product structure of the n -qubit Hilbert space, as each binary string directly corresponds to the tensor product of individual qubit states.

2 Decimal Representation

There is also the decimal representation which uses decimal numbers ranging from 0 to $N - 1$ (where $N = 2^n$) to label the basis states:

$$|0\rangle, |1\rangle, |2\rangle, \dots, |N - 1\rangle, \quad (6.50)$$

and a general n -qubit state can be written as

$$|\psi\rangle = \sum_{k=0}^{N-1} c_k |k\rangle. \quad (6.51)$$

In this representation, each basis state is labeled by a decimal number, which corresponds to the decimal equivalent of the binary string representation. The total number of basis states is still N . For example, for a 2-qubit system, the basis states are $|0\rangle$, $|1\rangle$, $|2\rangle$, and $|3\rangle$, which correspond to the binary notations $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$, respectively.



For those who are reading about this for the first time, it's important to note that the contents inside the $|\cdot\rangle$ are merely labels. In the computational basis, each basis state is a vector of length $N = 2^n$, having exactly one '1' in its structure, with the rest being zeros. See Eq. 6.10 for an example of two qubits.

This decimal representation is convenient when working with the Hilbert space as a whole, especially when dealing with quantum gates or other linear operators that act on the entire Hilbert space rather than individual qubits.

6.3.3 Product States, Entangled States, and Correlated States

We have already introduced the concepts of product and entangled states within the context of two-qubit systems. In the case of composite quantum systems comprising n -qubits, quantum states can be broadly classified into pure states and mixed states. Pure states include product states and entangled states. Mixed states can be correlated or uncorrelated.

1 Product States

A product state for an n -qubit composite quantum system is a state that can be expressed as the tensor product of the individual states of each qubit. Mathematically, an n -qubit product state can be written as:

$$|\psi_1\psi_2 \cdots \psi_n\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle, \quad (6.52)$$

where $|\psi_i\rangle$ signifies the state of the i -th qubit. In this scenario, the qubits are not entangled, meaning their states can be described independently of each other. Product states are separable, indicating that they can be factored into individual qubit states.

There can also be product states where only some qubits are independent of the rest, such as in the case of $|\psi_1\rangle \otimes |\psi_2 \cdots \psi_n\rangle$.

Verifying Properties of Tensor Products

To confirm the mathematical consistency of product states and operations involving them, consider the following properties of tensor products:

Due to the following property of tensor product,

$$(A_1 \otimes A_2 \cdots \otimes A_n)(B_1 \otimes B_2 \cdots \otimes B_n) = (A_1 B_1) \otimes (A_2 B_2) \cdots \otimes (A_n B_n), \quad (6.53)$$

we know the composite state is normalized if the component states are, because

$$\langle \psi_1\psi_2 \cdots \psi_n | \psi_1\psi_2 \cdots \psi_n \rangle = \langle \psi_1 | \psi_1 \rangle \otimes \langle \psi_2 | \psi_2 \rangle \cdots \otimes \langle \psi_n | \psi_n \rangle = 1. \quad (6.54)$$

Furthermore, operators can be 'distributed' into the tensor products:

$$\langle A_1 \otimes A_2 \cdots \otimes A_n | \psi_1 \otimes \psi_2 \cdots \otimes \psi_n \rangle = (A_1 | \psi_1 \rangle) \otimes (A_2 | \psi_2 \rangle) \cdots \otimes (A_n | \psi_n \rangle). \quad (6.55)$$

2 Entangled States

Entangled states are characterized by the inability to describe the quantum state of one qubit independently of the other qubits in the system. Entangled states exhibit stronger correlations than classical correlations, reflecting their inherently quantum mechanical nature. These states cannot be expressed as a product of individual qubit states and display nonlocal correlations that are fundamental to quantum information processing, including quantum communication, quantum cryptography, and quantum computing.

A prominent example of an entangled state is the GHZ state (Greenberger-Horne-Zeilinger state):

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|00\cdots 0\rangle + |11\cdots 1\rangle). \quad (6.56)$$

This state is non-factorizable into a product of individual qubit states, and the measurement outcomes on one qubit demonstrate nonlocal correlations with the measurement outcomes on the other qubits.

As an aside, let's use the following "compact" expressions of the GHZ state to illustrate some interesting notations associated with the tensor product:

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}} \left(\bigotimes_{i=1}^n |0\rangle + \bigotimes_{i=1}^n |1\rangle \right) \quad (6.57a)$$

$$= \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes n} + |1\rangle^{\otimes n} \right). \quad (6.57b)$$

These expressions illustrate the use of the tensor product symbol, \otimes , in the context of a state space composed of multiple qubits. The use of $\bigotimes_{i=1}^n$ and $^{\otimes n}$ effectively conveys the tensor product of n qubits, each in the state $|0\rangle$ or $|1\rangle$.

Exercise 6.7 Consider an n -qubit GHZ state given by Eq. 6.57, where $n > 2$. Determine the state of the rest of the system after the following measurements:

- (a) The first qubit is measured with an outcome $|0\rangle$.
- (b) The first qubit is measured with an outcome $|+\rangle$, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.
- (c) The first two qubits are measured with an outcome corresponding to the projection onto the state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.
- (d) The first two qubits are measured with an outcome corresponding to the projection onto the state $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$.

3 *Correlated States

Correlated states describe composite system states wherein the properties of one qubit are statistically linked to the properties of other qubits in the system. These correlations can arise due to interactions between the qubits, shared environments, or common preparation procedures. Correlated states can be either entangled or classically correlated.

Classically correlated states, while not entangled, still exhibit correlations between the qubits. These correlations are akin to classical correlations in statistical systems, wherein knowledge of one qubit provides information about other qubits, but each qubit can still be described by individual quantum states.

An example of a classically correlated n -qubit state is (expressed as a density matrix):

$$\rho_{1\dots n} = p_1 |00\dots 0\rangle\langle 0\dots 0| + p_2 |11\dots 1\rangle\langle 1\dots 1|, \quad (6.58)$$

where p_1 and p_2 are probabilities that satisfy $p_1 + p_2 = 1$.

Correlated states are a special case of mixed states, which we will explore in § 12.2.

6.3.4 *Identical Particles



While the discussion on fermions and bosons is not broadly applicable to general quantum computing, it does hold relevance in specialized scenarios such as quantum chemistry simulations. Furthermore, anyons directly link to emerging platforms in quantum computing. This section is thus intended to provide both context and an overview of particle statistics in quantum systems.

The Composition Postulate in quantum mechanics, as previously described, applies to distinguishable quantum systems. In these systems, like qubits, each unit can be uniquely identified, and the joint state space is the tensor product of the individual state spaces.

For unbound, indistinguishable particles, the joint state space is still constructed using the tensor product of the individual state spaces. However, the overall state (wavefunction) of the combined system must satisfy certain symmetry requirements, which depend on whether the particles are fermions or bosons.

Fermions and bosons are elementary particles that follow different statistical behaviors due to their intrinsic properties. Fermions, particles with half-integer spin such as electrons, protons, and neutrons, follow Fermi-Dirac statistics. Bosons, which have integer spin, like photons and certain atomic nuclei, follow Bose-Einstein statistics.

In contrast, qubits, even if they are based on half or integer spin particles, are localized in space, making them distinguishable from one another. Hence, their combined quantum states are formed according to the regular Composition Postulate, without needing to account for the symmetrization (for bosons) or antisymmetrization (for fermions) rules that apply to free, indistinguishable particles.

When combining indistinguishable quantum systems, the Composition Postulate is modified as follows:

- Fermions: Fermions adhere to the Pauli Exclusion Principle, which prevents two fermions from simultaneously occupying the same quantum state. Therefore, the combined state of multiple identical fermions must be antisymmetric under the exchange of any pair of fermions. This can be mathematically expressed using a Slater determinant or through second quantization formalism, which

ensures that the overall wavefunction changes sign upon the exchange of any two fermions.

- **Bosons:** Bosons, unlike fermions, can simultaneously occupy the same quantum state. The combined state of multiple identical bosons must be symmetric under the exchange of any pair of bosons. This symmetry can be expressed using symmetrization operators or second quantization formalism, which guarantees that the overall wavefunction remains unchanged upon the exchange of any two bosons.
- **Anyons:** Anyons are a class of quasiparticles in two-dimensional systems that do not strictly follow Fermi-Dirac or Bose-Einstein statistics. Instead, their wave functions acquire a complex phase upon exchanging positions, providing a unique behavior that lies between fermionic and bosonic statistics. This property makes them particularly intriguing for quantum computing. Anyonic systems, such as those realized in certain quantum Hall systems, are actively researched as platforms for topological quantum computing. In this computational paradigm, the quantum information is stored in the topological features of the system, making it robust against local errors. The anyonic braiding operations, where anyons are moved around each other, can serve as fault-tolerant quantum gates, significantly contributing to the robustness and scalability of a quantum computer.

6.4 * Measurements of Multi-Qubit Systems

The principles developed through two-qubit systems extend to multi-qubit systems for measurements, as outlined in § 6.2. We now enrich this understanding with additional insights and generalized formulae.

6.4.1 * Measurements on a Single Qubit

Assume we have an n -qubit system in a general superposition state:

$$|\psi\rangle = \sum_{x_1, x_2, \dots, x_n \in \{0,1\}} c_{x_1 x_2 \dots x_n} |x_1 x_2 \dots x_n\rangle. \quad (\text{Copy of 6.48})$$

If we measure the first qubit in the computational basis, the post-measurement state of the entire system will exist as a superposition of states where the first qubit is either in state $|0\rangle$ or in state $|1\rangle$, depending on the measurement outcome.

Assuming $|0\rangle$ is the measurement result, the associated probability is

$$P_0^{(1)} = \sum_{x_2, \dots, x_n \in \{0,1\}} |c_{0x_2 \dots x_n}|^2. \quad (6.59)$$

The post-measurement state of the entire system would comprise a superposition of all states with the first qubit in state $|0\rangle$, normalized by the probability of obtaining $|0\rangle$:

$$|\psi'\rangle = \frac{1}{\sqrt{P_0^{(1)}}} \sum_{x_2, \dots, x_n \in \{0,1\}} c_{0x_2 \dots x_n} |0x_2 \dots x_n\rangle \quad (6.60a)$$

$$= |0\rangle \otimes |\psi'_{n-1}\rangle. \quad (6.60b)$$

After the measurement, the first qubit becomes disentangled from the rest of the system. Here $|\psi'_{n-1}\rangle$ represents the post-measurement state of the remaining $n - 1$ qubits:

$$|\psi'_{n-1}\rangle = \frac{1}{\sqrt{P_0^{(1)}}} \sum_{x_2, \dots, x_n \in \{0,1\}} c_{0x_2 \dots x_n} |x_2 \dots x_n\rangle. \quad (6.61)$$

Let's introduce the *partial product notation*:

$$\langle y^{(j)} | x_1 x_2 \dots x_j \dots x_n \rangle \equiv \langle y | x_j \rangle |x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_n\rangle. \quad (6.62)$$

Note that in this notation, the bra $\langle y^{(j)} |$ does not match in length with the ket $|x_1 x_2 \dots x_n\rangle$. Instead, the index j in $\langle y^{(j)} |$ indicates that it interacts only with the corresponding $|x_j\rangle$ through the inner product, while leaving the states of all other qubits unchanged. This concept can be generalized to situations where there are multiple y 's.

Then the probability of measuring the first qubit with outcome $|0\rangle$ can be expressed in this shorthand form:

$$P_0^{(1)} = \left\| \langle 0^{(1)} | \psi \rangle \right\|^2. \quad (6.63)$$

And the post-measurement state $|\psi'_{n-1}\rangle$ is given by:

$$|\psi'_{n-1}\rangle = \frac{1}{\sqrt{P_0^{(1)}}} \langle 0^{(1)} | \psi \rangle. \quad (6.64)$$

An analogous treatment can be done for the case where we obtained $|1\rangle$ as the measurement result, but it won't be elaborated here.

Exercise 6.8 Consider the methodology outlined in Eqs. 6.60 to 6.64 for the case of measuring the first qubit and obtaining the outcome $|0\rangle$.

Your task is to replicate this analysis, but now assume that it's the last qubit being measured, and the measurement outcome is $|1\rangle$.

6.4.2 * Measurements on a Group of Qubits

To extend our previous discussion, where only the first qubit was measured, let's consider dividing the system into two sections, each referred to as a 'register': one with m qubits and the other with $k \equiv n - m$ qubits. If we perform a measurement on the m -qubit register, the post-measurement state of the entire system, depending on the measurement outcome, will be a superposition of states corresponding to that outcome, normalized appropriately. This process is often referred to as *partial measurement*.

For instance, the probability of obtaining a certain outcome $|a_1 a_2 \dots a_m\rangle$ for measuring the first m qubits is given by, in the partial product notation:

$$P_m = \left\| \langle a_1^{(1)} a_2^{(2)} \dots a_m^{(m)} | \psi \rangle \right\|^2 \quad (6.65a)$$

$$= \sum_{x_{m+1}, \dots, x_n \in \{0,1\}} |c_{a_1 a_2 \dots a_m x_{m+1} \dots x_n}|^2. \quad (6.65b)$$

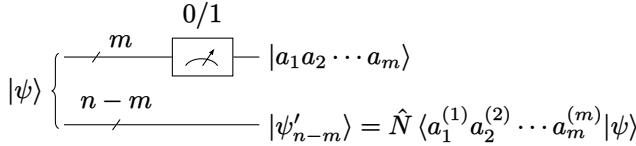
And the post-measurement state of the remaining $n - m$ qubits is:

$$|\psi'_{n-m}\rangle = \frac{1}{\sqrt{P_m}} \langle a_1^{(1)} a_2^{(2)} \cdots a_m^{(m)} | \psi \rangle \quad (6.66a)$$

$$= \frac{1}{\sqrt{P_m}} \sum_{x_{m+1}, \dots, x_n \in \{0,1\}} c_{a_1 a_2 \cdots a_m x_{m+1} \cdots x_n} |x_{m+1} \cdots x_n\rangle. \quad (6.66b)$$

The state $|\psi'_{n-m}\rangle$ resides within a subspace of the complete Hilbert space, spanned by the basis states $\{|x_{m+1} \cdots x_n\rangle\}$. The overall system state is given by:

$$|\psi'\rangle = |a_1 a_2 \cdots a_m\rangle \otimes |\psi'_{n-m}\rangle. \quad (6.67)$$



Notations: $a_1 a_2 \cdots a_m \in \{0, 1\}$ represents a series of binary values, \hat{N} denotes normalization, and $\langle a_1^{(1)} a_2^{(2)} \cdots a_m^{(m)} | \psi \rangle$ makes use of the partial product notation as defined in Eq. 6.62.

Figure 6.3: Measurement on a Group of Qubits

6.4.3 *A Partial Measurement Paradigm for Quantum Algorithms

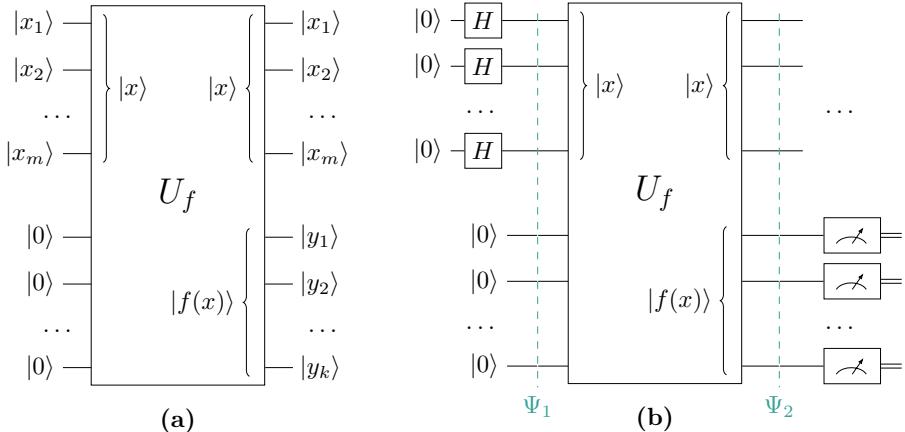


Figure 6.4: A Partial Measurement Paradigm for Quantum Algorithms

As an application of the partial measurement formulation discussed above, we examine a foundational model of measurement frequently employed in quantum algorithms.

1 Implementation of a Function

In Fig. 6.4(a), we show how a function $f(x)$ is commonly implemented as a unitary transformation U_f . The first m -qubit register, labeled $|x_1\rangle, |x_2\rangle, \dots, |x_m\rangle$, encodes

the input x . You can think of x as an m -bit string or a number made of the m binary digits.

The output of the function is the second register of k qubits, labeled $|y_1\rangle, |y_2\rangle, \dots, |y_k\rangle$. Similarly, you can regard these as representing the binary digits of the numerical output of $f(x)$. The function can also be understood in vector form:

$$[y_1, y_2, \dots, y_k]^T = f([x_1, x_2, \dots, x_m]^T). \quad (6.68)$$

The function f can be a one-to-one or many-to-one function. Below are some examples:

1. IsPrime Function: The output is one bit. $y = 1$ if x is a prime number, and $y = 0$ otherwise.
2. Mod-N Function: The output needs at least $\log_2 N$ qubits, representing $y = x \bmod N$. Here, N is an integer smaller than 2^k .

Note that the output is put into a separate second register, while the input is preserved in the first register. It is designed this way so that the transformation is U_f invertible (in fact, unitary), as required by quantum mechanics.

2 Uniform Superposition Input

Employing a superposition state as the input is a powerful tool of quantum computation, as demonstrated in Fig. 6.4(b). The Hadamard gate, denoted by H , transforms $|0\rangle$ into $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. This operation equips the input register at Ψ_1 with a uniform superposition of *all m-qubit basis states*, which is mathematically formulated as:

$$|\text{in}\rangle = |+\rangle^{\otimes m} = \sum_{x \in \{0,1\}^m} |x\rangle, \quad (6.69)$$

and the combined system state at the input is given by:

$$|\Psi_1\rangle = \sum_{x \in \{0,1\}^m} |x\rangle \otimes |0\rangle^{\otimes k}. \quad (6.70)$$

Due to the linearity of the transform U_f , the output register becomes:

$$|\text{out}\rangle = \sum_{x \in \{0,1\}^m} |f(x)\rangle, \quad (6.71)$$

and the combined system state at the output is given by:

$$|\Psi_2\rangle = \sum_{x \in \{0,1\}^m} |x\rangle \otimes |f(x)\rangle. \quad (6.72)$$

3 Measuring the Output

When we measure the output register, we obtain one of the possible values of $f(x)$, which we denote as \tilde{y} . Note that each measurement yields a \tilde{y} randomly. According to Eq. 6.66, the system collapses into a state where the second register is $|\tilde{y}\rangle$. Simultaneously, the first register is projected onto a superposition of only

those states $|x\rangle$ for which $f(x) = \tilde{y}$. This effectively singles out, within the first register, the superposition of all states corresponding to the pre-image of \tilde{y} under f . This superposition, which mathematically represents the solution set to the equation $f(x) = \tilde{y}$, is denoted by $|\tilde{x}\rangle$:

$$|\tilde{x}\rangle = \sum_{\substack{x \in \{0,1\}^m \\ f(x) = \tilde{y}}} |x\rangle. \quad (6.73)$$

Below are some examples:

1. IsPrime: If $f(x)$ is the IsPrime function given above, and the measurement outcome is 1, then $|\tilde{x}\rangle$ contains all the prime numbers less than 2^m . Conversely, if the measurement outcome is 0, $|\tilde{x}\rangle$ contains all the non-prime numbers less than 2^m .
2. Modular Exponentiation: $f(x) = p^x \bmod N$, where N and p are coprime integers with $p < N$. This function is periodic with some period r , such that $f(x) = f(x + nr)$ for any integer n .

Upon measuring the output and obtaining a value \tilde{y} , if $f(x_0) = \tilde{y}$ for some x_0 , then $|\tilde{x}\rangle$ is a superposition of the form $|x_0\rangle + |x_0 + r\rangle + |x_0 + 2r\rangle + |x_0 + 3r\rangle + \dots$, due to the periodicity of $f(x)$. Applying the quantum Fourier transform to $|\tilde{x}\rangle$ can reveal the period r . Discovering this period r enables an efficient method to factorize N , which is the foundational principle of Shor's algorithm.

This measurement paradigm is a key component in many quantum algorithms, including those of Simon and Shor, as well as in schemes for public-key quantum money (§ 11.3.2).

Exercise 6.9 Consider the Mod-N function with $N = 5$ and $m = 5$, i.e., $y = x \bmod 5$ with $0 \leq x < 32$. Assuming you have already prepared the system in a state of uniform superposition over all possible x values, apply the partial measurement paradigm as detailed in § 6.4.3.

- (a) What are the possible measurement outcomes?
- (b) If the measurement outcome is $\tilde{y} = 1$, what is the set of values for \tilde{x} ?

6.4.4 * Measurements in Alternate Bases

While most quantum computing devices natively support measurements in the computational basis, there are scenarios demanding measurements in other bases. In such cases, one would typically apply a basis rotation just before the measurement.

Consider an n -qubit state, $|\psi\rangle$. If we wish to measure it in an orthonormal basis, denoted by $\{|\phi_i\rangle\}$, the necessary unitary transformation U can be constructed as:

$$U = \sum_{i=0}^{2^n-1} |i\rangle\langle\phi_i|, \quad (6.74)$$

where $|i\rangle$ refers to an n -qubit computational basis state in decimal notation.

As demonstrated in § 3.4.5, a measurement of $U|\psi\rangle$ in the computational basis is equivalent to measuring $|\psi\rangle$ in the basis $\{|\phi_i\rangle\}$. This concept is illustrated in Fig. 6.5.

The probabilities from the measurement form a distribution over the possible outcomes, which represents the likelihood of the state $|\psi\rangle$ being projected onto each of the basis states, $\{|\phi_i\rangle\}$:

$$P_i = |\langle \phi_i | \psi \rangle|^2. \quad (6.75)$$

An equivalent perspective is to consider the probabilities associated with $U|\psi\rangle$ with respect to the computational basis:

$$P_i = |\langle i | U | \psi \rangle|^2. \quad (6.76)$$

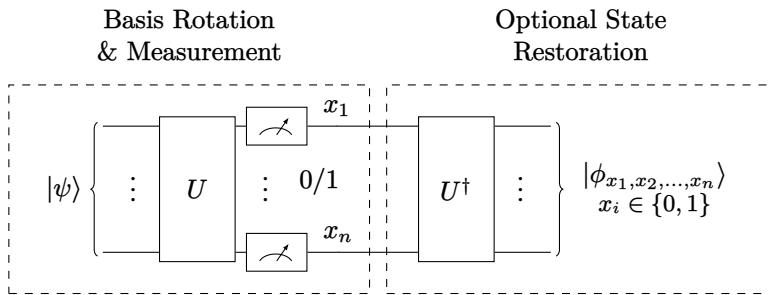
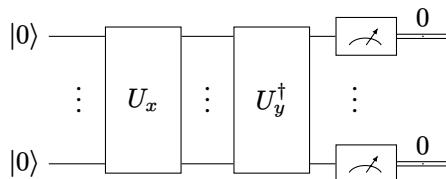


Figure 6.5: Multi-Qubit Measurement in Alternate Basis

Exercise 6.10 In many quantum machine learning algorithms, a specific circuit structure is observed. The depiction below showcases this typical circuit. Notably, the measurement is conducted against the $|00\dots 0\rangle$ state.



Consider the scenario where this measurement is performed repeatedly on the circuit. After accumulating a significant number of repetitions, one can derive a statistical average of the readouts.

Given this context, your tasks are as follows:

- Determine the expression for the measurement result, articulating it in terms of an inner product.
- Discuss the interpretation of the derived measurement result.

6.5 * Time Evolution of Multi-Qubit States

We have learned that Schrödinger equation governs the time evolution of the state of a quantum system, such as a qubit or an atom. In the context of multi-qubit

systems, the Schrödinger equation governs the time evolution of the joint state vector of the system, given by Eq. 6.48.

6.5.1 *Two-Qubit System

The Hamiltonian of a system of two qubits, A and B , is composed of individual Hamiltonians for each qubit and a possible interaction term:

$$H_{AB} = H_A \otimes I_B + I_A \otimes H_B + H_{\text{int}}. \quad (6.77)$$

Here, H_A and H_B are the Hamiltonians of qubits A and B , and I_A and I_B are the identity operators in their respective Hilbert spaces. The term H_{int} takes into account of the interaction between the two qubits, which depends on the specific physical implementation of the qubits and the type of interaction between them.

The Schrödinger equation for the two-qubit system is given by:

$$i\hbar \frac{\partial}{\partial t} |\psi_{AB}(t)\rangle = H_{AB} |\psi_{AB}(t)\rangle. \quad (6.78)$$

This equation governs the time evolution of the joint state vector $|\psi_{AB}(t)\rangle$ in the presence of interactions between the two qubits.

The solution to this time-dependent Schrödinger equation can be expressed formally in terms of a time-ordered exponential:

$$|\psi_{AB}(t)\rangle = e^{-iH_{AB}t/\hbar} |\psi_{AB}(0)\rangle. \quad (6.79)$$

This is a direct application of the time-evolution operator. Note that this solution assumes that the Hamiltonian does not explicitly depend on time.

6.5.2 *Multi-Qubit System

For an n -qubit system, the Schrödinger equation and Hamiltonian generalize naturally. The Hamiltonian of an n -qubit system is composed of individual Hamiltonians for each qubit and all possible interaction terms:

$$H = \sum_{i=1}^n H_i \otimes I_{-i} + \sum_{i < j} H_{ij} \otimes I_{-ij} + \dots. \quad (6.80)$$

Here, H_i are the individual Hamiltonians associated with the i -th qubit. These Hamiltonians govern the local dynamics of each qubit, which could be due to an external field or some local potential. I_{-i} denotes the identity operators on all other qubits except the i -th qubit. By using the identity operator on these other qubits, we ensure that the Hamiltonian H_i only acts on the i -th qubit and leaves the other qubits unaffected.

H_{ij} are the interaction Hamiltonians between the i -th and j -th qubits. These represent the influence of one qubit on another and could describe a variety of interactions, such as spin-spin interactions or couplings mediated by some external field. Again, I_{-ij} are the identity operators on all other qubits except the i, j -th qubits, ensuring that the interaction Hamiltonian H_{ij} only affects the qubits it is meant to.

Note that this is a general form of a multi-qubit Hamiltonian and the exact form of H_i and H_{ij} will depend on the specific quantum system being described. For example, in the Ising model discussed below, the H_{ij} terms are given by $J_{ij}Z_iZ_j$.

The ellipsis at the end of the equation indicates that this structure can be extended to incorporate interactions among larger groups of qubits, should your specific system require it. However, in many practical cases, one often restricts to single-qubit and two-qubit interactions due to their physical realisability and computational manageability.

The Schrödinger equation for the n -qubit system becomes:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle. \quad (6.81)$$

If the Hamiltonian does not explicitly depend on time, the above equation has a solution given by:

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle, \quad (6.82)$$

where $U(t)$ is the time-evolution operator:

$$U(t) = e^{-iHt/\hbar}. \quad (6.83)$$

6.5.3 * Example: The Ising Model

The transverse field Ising model (TFIM), Ising model for short, is a well-known model in condensed matter physics that describes a system of spins under the influence of a transverse magnetic field. In the context of quantum computing, it serves as a canonical example of a multi-qubit system whose ground state can be efficiently prepared on a quantum computer, and therefore, it can be used for various quantum optimization and simulation algorithms.

The Hamiltonian of the TFIM is given by:

$$H = \sum_{i < j} J_{ij} Z_i Z_j - \Gamma \sum_i X_i. \quad (6.84)$$

There are two terms that make up the Hamiltonian of the transverse field Ising model:

(1) $\sum_{i < j} J_{ij} Z_i Z_j$: This is the Ising interaction term. Here the sum is taken over all pairs of neighboring spins i and j on a lattice. The $i < j$ in the sum ensures the pairs are not double counted. J_{ij} denotes the strength of interaction between the i^{th} and j^{th} spins. Z_i and Z_j are the Pauli Z operators that operate on the spins at the i^{th} and j^{th} locations, respectively. Thus, $Z_i Z_j$ represents the operation performed on the spins at the i^{th} and j^{th} locations, i.e., it operates on the state $|i\rangle |j\rangle$. See Fig. 6.6.

i

In the above notation, $Z_i Z_j$ is a shorthand for $Z_i \otimes Z_j$, acting on spins i and j .

Similarly, X_i is a shorthand for $X_i \otimes I_{-i}$, where I_{-i} denotes the identity operators on all other qubits except the i -th qubit.

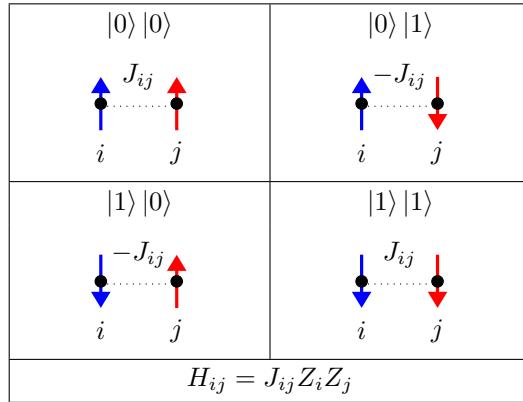


Figure 6.6: Illustration of Ising Interaction

(2) $-\Gamma \sum_i X_i$: This is the transverse field term, where the sum runs over all spins i on the lattice. The parameter Γ denotes the strength of the transverse magnetic field. The Pauli X operator, represented by X_i , operates on the spin at the i^{th} location and flips it. The negative sign indicates that the transverse field tries to align the spins along the x -axis, opposing the Ising interaction.

The two terms do not commute, making the exact time evolution of this system non-trivial. However, using methods such as Trotterization, one can efficiently simulate the time evolution of the TFIM on a quantum computer.

The Ising model provides a bridge between physical systems and computational problems, allowing the use of quantum devices to solve problems of practical interest. We will explore its applications in § 11.2.

6.5.4 *Example: CNOT Gate Implementation

The CNOT gate is a two-qubit operation. In many high-level descriptions, the intricacies of this operation are simplified using quantum gates depicted by their unitary transformations. However, for professionals engaged in quantum hardware implementation, comprehending the foundational concepts dictated by the Schrödinger equation remains essential.

The formal definition of the CNOT operator is provided in § 7.1.1:

$$\text{CNOT} = \begin{bmatrix} I & 0 \\ 0 & X \end{bmatrix} \quad (6.85a)$$

$$= \frac{I+Z}{2} \otimes I + \frac{I-Z}{2} \otimes X \quad (6.85b)$$

$$= \frac{1}{2}(I \otimes I + Z \otimes I + I \otimes X - Z \otimes X). \quad (6.85c)$$

In the context of Hamiltonian evolution, the key element in the CNOT operator is the term $\frac{1}{2}(Z \otimes X)$, symbolizing the interaction between the two qubits.

The typical Hamiltonian, instrumental in realizing the CNOT gate, is represented as:

$$H_{\text{CNOT}} = -\frac{\pi\hbar}{4\Delta t}(Z \otimes X). \quad (6.86)$$

Utilizing this Hamiltonian, we can substantiate that during a time evolution spanning Δt and accompanied by singular qubit rotations, the two-qubit system undergoes a transformation analogous to the CNOT gate action.

To validate this, let us revisit a generic matrix property: for any normal matrix A with $A^2 = I$, it holds that $e^{i\theta A} = \cos \theta I + i \sin \theta A$. In the context of our scenario, $(Z \otimes X)^2 = I \otimes I$, where $I \otimes I$ is the 4×4 identity matrix. Thus, we can infer:

$$U(\Delta t) = e^{-iH_{\text{CNOT}}\Delta t/\hbar} \quad (6.87a)$$

$$= e^{i\frac{\pi}{4}(Z \otimes X)} \quad (6.87b)$$

$$= \cos \frac{\pi}{4}(I \otimes I) + i \sin \frac{\pi}{4}(Z \otimes X) \quad (6.87c)$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i & 0 & 0 \\ i & 1 & 0 & 0 \\ 0 & 0 & 1 & -i \\ 0 & 0 & -i & 1 \end{bmatrix}. \quad (6.87d)$$

While the above matrix is not a precise replica of the CNOT operator, aligning it with the CNOT gate becomes feasible upon introducing a Z -rotation for the first qubit and an X -rotation for the second qubit:

$$\text{CNOT} = e^{i\frac{\pi}{4}} U(\Delta t) \left(R_z\left(\frac{\pi}{2}\right) \otimes R_x\left(\frac{\pi}{2}\right) \right), \quad (6.88)$$

where the rotation gates R_x and R_z are elaborated in § 5.2.5.

Exercise 6.11 Validate Eq. 6.88.

Hint: $U(\Delta t) = \begin{bmatrix} R_x(-\frac{\pi}{2}) & 0 \\ 0 & R_x(\frac{\pi}{2}) \end{bmatrix}$, $\text{CNOT} = U(\Delta t) \begin{bmatrix} R_x(\frac{\pi}{2}) & 0 \\ 0 & iR_x(\frac{\pi}{2}) \end{bmatrix}$.

Exercise 6.12 Illustrate that, if H_{CNOT} is directly represented as $\frac{\pi\hbar}{2\Delta t}$ CNOT (often deemed impractical), then the resulting $U(\Delta t) = -i$ CNOT.

Hint: $\text{CNOT}^2 = I$.

6.6 Summary and Conclusions

Conceptual Progression

The exploration of multi-qubit systems in this chapter serves as a cornerstone for more advanced topics in quantum computation. As we transitioned from the familiar realm of classical Newtonian mechanics with its additive principles, we dove into the quantum domain governed by the tensor product principle. This principle emerges as the defining aspect of quantum mechanics that distinguishes it from classical mechanics.

Our journey began with an examination of the two-qubit system, gradually expanding our scope to systems of n qubits. By doing so, we highlighted the

exponential complexity in terms of state space and computational capabilities. While practical applications remain limited due to challenges such as qubit decoherence, the theoretical underpinnings suggest immense potential, especially when error correction methods mature in the future.

Measurements in Multi-Qubit Systems

An integral part of this chapter was dedicated to understanding measurements in multi-qubit systems. We revisited the quantum measurement framework, extending its application to composite quantum systems, especially those with multiple qubits. The distinction between local measurements on subsystems and joint measurements was elaborated upon. Emphasis was placed on the tensor product structure of state space and the associated correlations between subsystems.

Time Evolution and Quantum Interactions

By invoking the Schrödinger equation, we shed light on the time evolution of multi-qubit systems. Exploring both two-qubit and n -qubit systems, we grasped the intricacies of their evolution. Practical applications and examples, such as the Ising model and the CNOT gate implementation, further illustrated these abstract principles.

Practical Implications and Challenges

While the theoretical implications of multi-qubit systems are profound, practical challenges such as qubit decoherence pose hurdles. This interplay between theory and practicality will be significant as quantum computing progresses, with error correction emerging as a fundamental area of research.

Upcoming Topics

As we venture into the subsequent chapters, our focus will shift to multi-qubit quantum gates. These gates, vital for quantum computing, operate on multiple qubits simultaneously and embody properties distinct from single-qubit gates. We will explore commonly used multi-qubit gates, universal gate sets, gate sequences, and introduce new tools like the Boolean representation of quantum gates and circuits. This transition will lay a firm foundation for advanced topics in quantum computing and information processing.

Problem Set 6

- 6.1** Consider $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, $|\alpha\rangle = \cos \alpha |0\rangle + \sin \alpha |1\rangle$, and $|\beta\rangle = \cos \beta |0\rangle + \sin \beta |1\rangle$, where $\alpha, \beta \in \mathbb{R}$.

Compute the probability $P(\alpha, \beta) = |\langle \alpha | \langle \beta | \psi \rangle|^2$.

- 6.2** Determine if the state $c(|00\rangle - 2i|01\rangle + 2i|10\rangle + 4|11\rangle)$ is a product state. If it is, identify the two component states.

- 6.3** (a) Define $\Pi_a = \frac{1}{2}(I - X)$ and $\Pi_b = \frac{1}{2}(I - Z)$. Show that Π_a and Π_b are projection operators, i.e., $\Pi^2 = \Pi$. Here, X and Z are Pauli operators.

- (b) Given $|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, calculate $\langle\psi|\Pi_a \otimes \Pi_b|\psi\rangle$.

6.4 Find a 4×4 Hamiltonian H such that e^{-iH} equals the CNOT matrix.

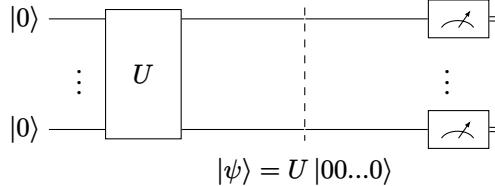
6.5 Consider a general n -qubit system in a superposition state as defined in Eq. 6.48. We will conduct a measurement on the first two qubits and obtain the outcome corresponding to one of the Bell states, specifically $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$.

Your task is to construct equations that determine the probability of this measurement outcome and the resulting post-measurement state of the system. You should follow the process outlined in Eqs. 6.66 and 6.67.

6.6 Consider a function $y = f(x)$ with a pairing property that $f(x) = f(x \oplus s)$ for any x , where \oplus denotes the bitwise XOR operation, and x , y , and s are all m -bit binary strings. This function is a component of Simon's algorithm.

Assuming you have already prepared the system in a state of uniform superposition over all possible x values, apply the partial measurement paradigm as detailed in § 6.4.3. Describe the resulting state $|\tilde{x}\rangle$ upon measurement of the system. (The Simon's algorithm hinges on getting information of s from $|\tilde{x}\rangle$ using a subsequent Hadamard transform.)

6.7 The following quantum circuit is run repeatedly for N times. For each run, each of the n qubits is measured in the computational basis; the k -th run yields n measurement values $z_1^{(k)}, z_2^{(k)}, \dots, z_n^{(k)}$, where $z_i^{(k)} = 1$ for measurement outcome $|0\rangle$, and $z_i^{(k)} = -1$ for $|1\rangle$. Essentially, we are measuring each qubit with the Z observable.



From the measurement results, do the following:

- Calculate the empirical probability of measuring the i -th qubit and obtaining $|0\rangle$. Estimate the sampling error.
- Calculate the empirical probability of measuring all n qubits and obtaining $|00\dots0\rangle$. Estimate the sampling error.
- Calculate the empirical expectation value $\langle Z_i \rangle$ for the i -th qubit. Estimate the sampling error.
- Calculate the empirical expectation value $\langle Z_i \otimes Z_j \rangle$ for the i -th and j -th qubits. Estimate the sampling error.
- Can you obtain the empirical expectation values $\langle X_i \rangle$ and $\langle Z_i \otimes X_j \rangle$? Explain. If not, how would you revise the circuit to obtain these values?

6.8 The single-qubit Hadamard transform is the unitary transformation defined by:

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

The n -qubit Hadamard transform is defined as the tensor product of n single-qubit Hadamard transforms:

$$H_n = H^{\otimes n} = H \otimes H \otimes \cdots \otimes H \quad (\text{\(n\) times}).$$

- (a) Write out the matrix form of H_3 .
- (b) Prove that H_n is a unitary transformation.
- (c) Calculate the state vector $H_n |00\cdots 0\rangle$.
- (d) Calculate the state vector $H_n |11\cdots 1\rangle$.
- (e) *Derive a general formula for $H_n |x_1 x_2 \cdots x_n\rangle$.
- (f) *Derive a general formula for $H_n \sum_{x_1, x_2, \dots, x_n \in \{0,1\}} c_{x_1 x_2 \cdots x_n} |x_1 x_2 \cdots x_n\rangle$.

- 6.9** The n -qubit XOR operation is a transformation often used in quantum algorithms such as Simon's algorithm. Given a binary string $s = a_1 a_2 \cdots a_n$ (for example $0101 \cdots$), the XOR transform is defined as follows:

$$\text{XOR}_n |x_1 x_2 \cdots x_n\rangle = |y_1 y_2 \cdots y_n\rangle,$$

where $y_j = x_j \oplus a_j$, for all j from 1 to n .

- (a) Prove that XOR_n is a unitary transformation.
- (b) Calculate the state vector $\text{XOR}_n |00\cdots 0\rangle$.
- (c) Calculate the state vector $\text{XOR}_n |11\cdots 1\rangle$.
- (d) *Derive a general formula for $\text{XOR}_n |x_1 x_2 \cdots x_n\rangle$.
- (e) *Derive a general formula for $\text{XOR}_n \sum_{x_1, x_2, \dots, x_n \in \{0,1\}} c_{x_1 x_2 \cdots x_n} |x_1 x_2 \cdots x_n\rangle$.
- (f) *Write out the matrix form of XOR_n .

- 6.10** The N -dimensional quantum Fourier transform (QFT) is given by

$$\text{QFT} |x\rangle = |\tilde{x}\rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \omega_N^{kx} |k\rangle, \quad \text{where } x = 0, 1, 2, \dots, N-1. \quad (6.89)$$

It can be viewed as a change from the computational basis $\{|x\rangle\}$ to the Fourier basis $\{|\tilde{x}\rangle\}$. Here $\omega_N = e^{\frac{2\pi i}{N}}$, $i = \sqrt{-1}$, and $\{\omega_N^k\}$ are the N -th complex roots of 1.

- (a) Find the basis transformation matrix QFT for $N = 4$.
- (b) Find the matrix form of QFT and QFT^\dagger for any N .
- (c) *Calculate the representation of $|\psi\rangle = \sum_k c_k |k\rangle$ in the Fourier basis.
- (d) *Compute $\langle A \rangle = \langle \psi | A | \psi \rangle$ in both the computational and Fourier bases and demonstrate they are equal, for $N = 4$ and $A = \sum_{k=0}^3 (k-2) |k\rangle \langle k|$.

7. Multi-Qubit Quantum Gates

Contents

7.1	Common Multi-Qubit Gates	171
7.1.1	CNOT Gate	173
7.1.2	Flipped CNOT Gate	174
7.1.3	* General Controlled-U Gate	175
7.1.4	Classically Controlled-U Gate	177
7.1.5	* Two-Qubit Rotation Gates	178
7.1.6	* Parallel Gates	178
7.2	Universal Sets of Qubit Gates	179
7.2.1	CNOT + Arbitrary Single-Qubit Rotations	180
7.2.2	Clifford Group + T	181
7.2.3	Toffoli + H	181
7.3	Boolean Representation of Quantum Gates	182
7.3.1	Properties of XOR Operations	182
7.3.2	Boolean Representation of Common Quantum Gates	183
7.3.3	Understanding Boolean Representation	184
7.4	Equivalent Gate Sequences	186
7.4.1	Order Independence in Separate Qubit Operations	187
7.4.2	Gate Sequence: A First Example	187
7.4.3	Sequences Involving X, Z, and CNOT	189
7.4.4	Sequences Involving H	190
7.4.5	* Deferred Measurement Principle	192
7.4.6	* Operator Sum Relations	194
7.5	* Exploratory Topics	195
7.5.1	* Expressing Common Gates with the Toffoli Gate	195
7.5.2	* Advanced Quantum Gate Examples	196
7.6	Summary and Conclusions	197
	Problem Set 7	198

Multi-qubit gates serve a critical function in quantum computing, primarily due to their ability to create and manipulate entangled states. These states, which are

essential for executing universal quantum algorithms, cannot be generated using single-qubit gates alone. Unlike single-qubit gates that apply unitary transformations to individual qubits, multi-qubit gates extend their action to multiple qubits simultaneously. They act on an n -qubit state space, making their corresponding matrix representation a $2^n \times 2^n$ matrix, where n denotes the number of qubits involved.

Despite the unique and critical role they play, multi-qubit gates maintain the key property of being unitary operators. They preserve the inner product between state vectors, ensuring that the laws of quantum mechanics, particularly the conservation of probability, remain unbroken after their application. This unitary nature makes them consistent with the overall framework of quantum mechanics and quantum computation.

This chapter starts by examining commonly used multi-qubit gates in quantum computing and universal gate sets. Building on these foundational concepts, we delve into the study of gate sequences, including a rigorous exploration of sequences that produce equivalent transformations on qubits. To aid this exploration, we introduce a novel tool: the Boolean representation of quantum gates and circuits. These topics are fundamental to the field of quantum computing and information processing and set the stage for further inquiry.

7.1 Common Multi-Qubit Gates

We will begin with an in-depth exploration of multi-qubit gates commonly used in quantum computing. Subsequently, we will delve into related topics such as parallel quantum gates and classically controlled gates.

Multi-qubit gates often encountered in quantum computing include the Controlled NOT (CNOT), Controlled Z (CZ), SWAP, Toffoli (also known as CCNOT), Fredkin (also known as CSWAP) gates, the general Controlled-U (CU) gate, and the two-qubit rotation gates. In the CU gate, U represents any single-qubit unitary transformation. These gates play essential roles in quantum circuits and quantum algorithms, enabling the creation of entangled states and execution of conditional operations based on the state of one or more control qubits.



The **highlighted** terms in the operators in Table 7.1 denote the changes in the basis mappings induced by the gates.

The matrix representation of the gates in Table 7.1 is provided in Eq. 7.1.

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & X \end{bmatrix}, \quad (7.1a)$$

$$\text{CZ} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix}, \quad (7.1b)$$

Gate	Symbol	Operator	Description
CNOT (CX)		$ 00\rangle\langle 00 + 01\rangle\langle 01 + 10\rangle\langle 11 + 11\rangle\langle 10 $	Applies X to the target qubit if the control qubit is $ 1\rangle$.
CZ (ZZ)		$ 00\rangle\langle 00 + 01\rangle\langle 01 + 10\rangle\langle 10 - 11\rangle\langle 11 $	Applies Z to the target qubit if the control qubit is $ 1\rangle$.
SWAP		$ 00\rangle\langle 00 + 01\rangle\langle 10 + 10\rangle\langle 01 + 11\rangle\langle 11 $	Exchanges the states of the two qubits it acts upon.
CCNOT (Toffoli)		$ 000\rangle\langle 000 + 001\rangle\langle 001 + 010\rangle\langle 010 + 011\rangle\langle 011 + 100\rangle\langle 100 + 101\rangle\langle 101 + 110\rangle\langle 111 + 111\rangle\langle 110 $	Applies X to the target qubit if both control qubits are $ 1\rangle$.
CSWAP (Fredkin)		$ 000\rangle\langle 000 + 001\rangle\langle 001 + 010\rangle\langle 010 + 011\rangle\langle 011 + 100\rangle\langle 100 + 101\rangle\langle 101 + 110\rangle\langle 110 + 111\rangle\langle 111 $	Swaps the states of the second and third qubits if the control qubit is $ 1\rangle$.
Ctrl-U (CU)		$ 00\rangle\langle 00 + 01\rangle\langle 01 + u_{00} 10\rangle\langle 10 + u_{01} 10\rangle\langle 11 + u_{10} 11\rangle\langle 10 + u_{11} 11\rangle\langle 11 $	Applies U to the target qubit if the control qubit is $ 1\rangle$.
$R_{zz}(\theta)$			Two-qubit rotation gates. Also, R_{zx} , R_{xy} , etc.

Note: The CZ gate exhibits a unique property: it does not matter whether qubit 1 or 2 serves as the control qubit (see Exercise 7.4). This symmetry in its behavior leads to its alternative name, the ZZ gate, which is represented by the symbol shown to the right.



Table 7.1: Common Multi-Qubit Quantum Gates

$$\text{CU} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & u_{00} & u_{01} \\ 0 & 0 & u_{10} & u_{11} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix}, \quad (7.1c)$$

$$\text{SWAP} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} I+Z & X-iY \\ X+iY & I-Z \end{bmatrix}, \quad (7.1d)$$

$$\text{CCNOT} = \begin{bmatrix} 8 \times 8 \\ \text{matrix} \end{bmatrix} = \begin{bmatrix} I_6 & 0 \\ 0 & X \end{bmatrix}, \quad (7.1e)$$

$$\text{CSWAP} = \begin{bmatrix} 8 \times 8 \\ \text{matrix} \end{bmatrix} = \begin{bmatrix} I_4 & 0 \\ 0 & \text{SWAP} \end{bmatrix}. \quad (7.1f)$$

Exercise 7.1 Verify that the operator matrices in Eq. 7.1 are unitary.

Exercise 7.2 Investigate why Controlled- H is not a commonly used quantum gate.

7.1.1 CNOT Gate

Let's use the CNOT gate, which is the most useful multi-qubit gate, to examine how its operator and matrix work. For the CNOT gate shown in Table 7.1, qubit 1 is the control qubit, and qubit 2 is the target qubit. When the control qubit is $|1\rangle$, an X (i.e., NOT) gate is applied to the target qubit, flipping its 0 and 1.

For example, the basis state $|10\rangle$ is transformed to $|11\rangle$. This corresponds to the term $|11\rangle\langle 10|$ in the operator of CNOT. For the CNOT matrix (see Eq. 7.2), we label the rows with $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$, and the columns with $\langle 00|$, $\langle 01|$, $\langle 10|$, and $\langle 11|$. Then $|11\rangle\langle 10|$ corresponds to the 1 at the row $|11\rangle$ and column $\langle 10|$. This is summarized as the third row in the following Table:

Basis Mapping	Projection Operator	Matrix Element
$ 00\rangle \mapsto 00\rangle$	$ 00\rangle\langle 00 $	1 at row $ 00\rangle$, col $\langle 00 $
$ 01\rangle \mapsto 01\rangle$	$ 01\rangle\langle 01 $	1 at row $ 01\rangle$, col $\langle 01 $
$ 10\rangle \mapsto 11\rangle$	$ 11\rangle\langle 10 $	1 at row $ 11\rangle$, col $\langle 10 $
$ 11\rangle \mapsto 10\rangle$	$ 10\rangle\langle 11 $	1 at row $ 10\rangle$, col $\langle 11 $

$$\text{CNOT} = \begin{bmatrix} |00\rangle & \langle 00| \\ |01\rangle & \langle 01| \\ |10\rangle & \langle 10| \\ |11\rangle & \langle 11| \end{bmatrix}. \quad (7.2)$$

The CNOT gate is idempotent, i.e., $\text{CNOT}^2 = I_4$, where $I_4 = I \otimes I$ is the 4×4 identity matrix. This means applying the CNOT gate twice in succession results in no change in the system.

A distinctive feature of quantum gates compared to classical logic gates is that quantum gates can work with superposition states. Let's examine a general two-qubit state:

$$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle. \quad (7.3)$$

The term $|11\rangle\langle 10|$ in the operator of CNOT picks out $\gamma|10\rangle$ from $|\psi\rangle$ and transforms it to $\gamma|11\rangle$. Similarly, for other terms in the operator. Thus

$$\text{CNOT } |\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|11\rangle + \delta|10\rangle. \quad (7.4)$$

The corresponding matrix operation is:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix}. \quad (7.5)$$

 The term “control qubit” in the context of the CNOT gate might suggest that the state of the control qubit remains invariant. However, this is not always true,

especially when the control qubit is in a superposition. Consider the $\{|+\rangle, |-\rangle\}$ basis. When applying the CNOT gate to the basis states, we observe:

$$\text{CNOT} |++\rangle = |++\rangle, \quad (7.6a)$$

$$\text{CNOT} |+-\rangle = |--\rangle, \quad (7.6b)$$

$$\text{CNOT} |-+\rangle = |-+\rangle, \quad (7.6c)$$

$$\text{CNOT} |--\rangle = |+-\rangle. \quad (7.6d)$$

It is evident from these results that, in the $\{|+\rangle, |-\rangle\}$ basis, the state of the “target qubit” remains unchanged, while the “control qubit” can change based on the operation. Therefore, *multi-qubit gates should always be regarded as operating on the all the qubits as an integrated system, rather than affecting them independently.*

Exercise 7.3 Verify Eq. 7.6.

7.1.2 Flipped CNOT Gate

Let's consider the CNOT gate with qubit 2 as control qubit and qubit 1 target qubit. We name it CNOT' or CX'. Now when the second qubit is $|1\rangle$, an X (i.e., NOT) gate is applied to the first qubit, flipping its 0 and 1.

Symbol, operator, and description:

Gate	Symbol	Operator	Description
CNOT' (CX')		$ 00\rangle\langle 00 + 11\rangle\langle 01 + 10\rangle\langle 10 + 01\rangle\langle 11 $	Applies X to the first qubit if the second qubit is $ 1\rangle$.

Basis mapping:

Basis Mapping	Projection Operator	Matrix Element
$ 00\rangle \mapsto 00\rangle$	$ 00\rangle\langle 00 $	1 at row $ 00\rangle$, col $\langle 00 $
$ 01\rangle \mapsto 11\rangle$	$ 11\rangle\langle 01 $	1 at row $ 11\rangle$, col $\langle 01 $
$ 10\rangle \mapsto 10\rangle$	$ 10\rangle\langle 10 $	1 at row $ 10\rangle$, col $\langle 10 $
$ 11\rangle \mapsto 01\rangle$	$ 01\rangle\langle 11 $	1 at row $ 01\rangle$, col $\langle 11 $

Matrix:

$$\text{CNOT}' = \begin{bmatrix} |00\rangle & |01\rangle & |10\rangle & |11\rangle \\ |00\rangle & 1 & 0 & 0 \\ |01\rangle & 0 & 0 & 0 \\ |10\rangle & 0 & 0 & 1 \\ |11\rangle & 0 & 1 & 0 \end{bmatrix}. \quad (7.7)$$

The CNOT' matrix can be written in 2×2 block form as:

$$\text{CNOT}' = \frac{1}{2} \begin{bmatrix} I + Z & I - Z \\ I - Z & I + Z \end{bmatrix}. \quad (7.8)$$

Exercise 7.4 Demonstrate that, in contrast to the fact that CNOT is not equal to CNOT', the gate CZ is equal to CZ'. Here, CZ' represents the CZ gate with qubit 2 serving as the control qubit and qubit 1 as the target qubit. In fact, because of this property, the CZ gate is also referred as the ZZ gate.

7.1.3 *General Controlled-U Gate

A Controlled-U (or CU) gate applies a generic unitary gate U to the target qubit if the control qubit is $|1\rangle$. The CNOT, CZ, CCNOT, and CSWAP gates listed in Table 7.1 are special cases of Controlled-U gates. The target and control can also be more than one qubit. For example, the CCNOT gate has two qubits for the control, and the CSWAP gate has two qubits for the target.

1 Mathematical Representations

Mathematically, the Controlled-U gate can be represented as:

$$\text{CU} = \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix}, \quad (7.9)$$

where U is a generic unitary matrix, and I is the identity matrix.

In the following, for simplicity, we only consider the case where the target and control are single qubits, so U and I are 2×2 matrices.

The above matrix for CU has a corresponding operator form:

$$\text{CU} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes U, \quad (7.10)$$

which, when expanded, yields the operator in Table 7.1.

The fact that the CU gate applies U to the target qubit when the control qubit is $|1\rangle$ can be expressed as:

$$\text{CU} |0\rangle |\psi\rangle = |0\rangle |\psi\rangle, \quad (7.11a)$$

$$\text{CU} |1\rangle |\psi\rangle = |1\rangle U |\psi\rangle. \quad (7.11b)$$

Exercise 7.5 Demonstrate that the representations of the CU gate are equivalent: Eq. 7.9, Eq. 7.10, Eq. 7.11, and the operator form in Table 7.1.

The skills involved in this exercise is foundational for working with quantum algorithms. Make sure you can do this exercise independently.

2 Properties

The Controlled-U gate has a number of important properties:

Unitarity

The Controlled-U gate is unitary if U is unitary.

Since Pauli operators are unitary, so are CNOT and CZ.

Proof. The matrix representation of CU is given by Eq. 7.9. The Hermitian conjugate of CU, denoted by CU^\dagger , is:

$$\text{CU}^\dagger = \begin{bmatrix} I & 0 \\ 0 & U^\dagger \end{bmatrix}. \quad (7.12)$$

Since U is unitary, $UU^\dagger = I$. Therefore,

$$\text{CU} \cdot \text{CU}^\dagger = \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & U^\dagger \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = I, \quad (7.13)$$

which proves that CU is unitary if U is unitary. \square

Not a Parallel Gate

The CU gate cannot be decomposed as a tensor product $W \otimes U$ where U and W are non-identity unitary matrices.

This property implies that the application of a Controlled-U gate involves interaction between qubits.

A gate with an operator of the form $W \otimes U$ is a parallel gate, which will be discussed in § 7.1.6.

Proof.

The matrix representation of CU is given by Eq. 7.9. Now, assume for the sake of contradiction that we could write CU as a product operator, $W \otimes U$. Then, this would imply that

$$W \otimes U = \begin{bmatrix} w_{00}U & w_{01}U \\ w_{10}U & w_{11}U \end{bmatrix}, \quad (7.14)$$

for some matrix W . Comparing this with the matrix of CU in Eq. 7.9, we have the following system of equations:

$$w_{00}U = I, \quad (7.15a)$$

$$w_{01}U = 0, \quad (7.15b)$$

$$w_{10}U = 0, \quad (7.15c)$$

$$w_{11}U = U. \quad (7.15d)$$

For $w_{01}U$ to be the zero matrix, w_{01} must be zero since U is a unitary matrix. Similarly, w_{10} must be zero. This leads us to

$$W \otimes U = \begin{bmatrix} w_{00}U & 0 \\ 0 & w_{11}U \end{bmatrix}. \quad (7.16)$$

Comparing this with the matrix of CU, we can see that this form can represent the CU gate only if $w_{00} = 1$, $w_{11} = 1$, and $U = I$.

Therefore, we conclude that the CU gate is not a simple product operator like $W \otimes U$. \square

Involutory Property

If $U^2 = I$, then $\text{CU}^2 = I_{2 \times 2}$.

Proof.

Given that $U^2 = I$, we can write

$$\text{CU}^2 = \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & U^2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = I_{2 \times 2}. \quad (7.17)$$

□

3 Controlled- U with Phase

For a standalone U gate, its global phase does not have any physical effect; that is, U and $e^{i\phi}U$ are equivalent as quantum gates. However, for a controlled- U gate, the global phase of U does matter, because

$$\begin{bmatrix} I & 0 \\ 0 & e^{i\phi}U \end{bmatrix} \neq e^{i\phi} \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix}. \quad (7.18)$$

In particular, CNOT and CZ have their “negative” counterparts, which are distinct from their original versions:

$$\overline{\text{CNOT}} = \begin{bmatrix} I & 0 \\ 0 & -X \end{bmatrix}, \quad (7.19a)$$

$$\overline{\text{CZ}} = \begin{bmatrix} I & 0 \\ 0 & -Z \end{bmatrix}. \quad (7.19b)$$

Exercise 7.6 Evaluate $\overline{\text{CNOT}}|++\rangle$, $\overline{\text{CNOT}}|+-\rangle$, $\overline{\text{CNOT}}|-+\rangle$, and $\overline{\text{CNOT}}|--\rangle$.

7.1.4 Classically Controlled-U Gate

Controlled gates like CNOT, CZ, and CU are quantum controlled gates that operate based on the states of control qubits. However, there also exist classically controlled gates, which perform actions dictated by classical logic. These gates involve the inclusion or exclusion of certain gates in the instruction sequence for the quantum computer based on classical control signals.

Classically controlled gates function similarly to their quantum controlled counterparts. However, there is a key difference. In a quantum controlled gate, the control qubit can be in a superposition state, whereas a classically controlled gate is driven by a classical bit (0 or 1), often determined by the measurement outcome of a qubit. The operator for a quantum controlled U gate is given by Eq. 7.10, whereas for a classically controlled U gate,

$$\text{cCU} = U^i = \begin{cases} I & \text{for } i = 0, \\ U & \text{for } i = 1, \end{cases} \quad (7.20)$$

where i denotes the control bit.

The symbol for a classically controlled gate employs a double line for the control wire, as shown in Fig. 7.1.

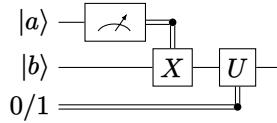


Figure 7.1: Examples of Classically Controlled Gates

7.1.5 *Two-Qubit Rotation Gates

Two-qubit rotation gates are essential components in quantum algorithms, facilitating complex interactions between qubits. These gates generalize the notion of single-qubit rotations and find applications in quantum algorithms like VQE and VAOA, which will be discussed in § 11.2. They can be represented as follows:

$$R_{pq}(\theta) = e^{-i\frac{\theta}{2}P \otimes Q}, \quad P, Q \in \{X, Y, Z\}. \quad (7.21)$$

Recall that for any normal matrix A with $A^2 = I$, it holds that $e^{i\theta A} = \cos \theta I + i \sin \theta A$. In the present case, $(P \otimes Q)^2 = I \otimes I$, where $I \otimes I$ is the 4×4 identity matrix. Thus, we can infer:

$$R_{pq}(\theta) = \cos \frac{\theta}{2}(I \otimes I) - i \sin \frac{\theta}{2}(P \otimes Q). \quad (7.22)$$

For example, the $R_{zz}(\theta)$ gate, also commonly termed the parameterized ZZ gate, is defined as:

$$R_{zz}(\theta) \equiv \text{ZZ}(\theta) = \cos \frac{\theta}{2}(I \otimes I) - i \sin \frac{\theta}{2}(Z \otimes Z) \quad (7.23a)$$

$$= \begin{bmatrix} R_z(\theta) & 0 \\ 0 & R_z(-\theta) \end{bmatrix} \quad (7.23b)$$

$$= \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 & 0 & 0 \\ 0 & e^{i\frac{\theta}{2}} & 0 & 0 \\ 0 & 0 & e^{i\frac{\theta}{2}} & 0 \\ 0 & 0 & 0 & e^{-i\frac{\theta}{2}} \end{bmatrix}. \quad (7.23c)$$

Exercise 7.7 Explain why $R_{zz}(\theta) \neq R_z(\theta) \otimes R_z(\theta)$.

$$R_{zx}(\theta) = \cos \frac{\theta}{2}(I \otimes I) - i \sin \frac{\theta}{2}(Z \otimes X) \quad (7.24a)$$

$$= \begin{bmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} & 0 & 0 \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} & 0 & 0 \\ 0 & 0 & \cos \frac{\theta}{2} & i \sin \frac{\theta}{2} \\ 0 & 0 & i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix}. \quad (7.24b)$$

7.1.6 *Parallel Gates

A gate with an operator of the form $W \otimes U$ is referred to as a parallel gate. When acting on unentangled qubits, W and U act in parallel on their respective qubits, behaving as two independent gates, as illustrated in Eq. 7.25 and Fig. 7.2. In

this case, if W or U equals the identity operator I , the parallel gate reduces to a single-qubit gate.

$$(W \otimes U)(|\psi\rangle \otimes |\phi\rangle) = (W|\psi\rangle) \otimes (U|\phi\rangle). \quad (7.25)$$

Equation 7.25 is applicable for product states, where $|\psi\rangle$ and $|\phi\rangle$ are unentangled.

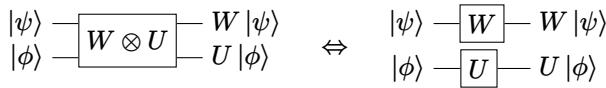


Figure 7.2: Parallel Gate on Unentangled State

Parallel gates are considered “trivial” multi-qubit gates, which is why the common multi-qubit gate discussed in § 7.1 has not included any of those. However, there is an important non-trivial application of parallel gates: when we want to apply a gate to part of an entangled state $|\psi\rangle$, for example, applying U to the second qubit of a Bell state, we are effectively applying the parallel gate $I \otimes U$ to the combined system, even though we may still draw a single qubit gate on the second qubit, as illustrated in Fig. 7.3.

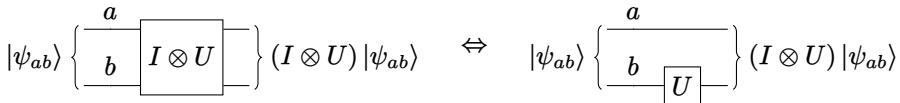


Figure 7.3: Gate on Part of an Entangled State

We often use a shorthand notation for $(I \otimes U)|\psi_{ab}\rangle$: $U_b|\psi_{ab}\rangle$, where U_b stands for U applied to qubit- b . This notation is convenient when we work with basis states, for example,

$$(I \otimes X) \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = X_2 \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \quad (7.26a)$$

$$= \frac{1}{\sqrt{2}}(|0\rangle X|0\rangle + |1\rangle X|1\rangle) \quad (7.26b)$$

$$= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle). \quad (7.26c)$$

7.2 Universal Sets of Qubit Gates

In classical computing, complex operations are implemented as sequences of simpler operations. A combination of the NAND gate and the FANOUT circuit element can construct digital circuits capable of performing all computational tasks, forming what is known as a universal set of gates for classical computing. However, the NAND and FANOUT gates are not reversible. If we restrict ourselves to reversible gates, achieving universality with only one- and two-bit gates becomes impossible. Notably, a set consisting solely of the Toffoli gate can serve as a universal gate set for classical reversible computation.

Quantum computing employs gates as unitary transformations, which leads to an infinite number of possible gates for single, two, and multi-qubit operations on a quantum computer. This is in stark contrast to classical computing, which only has four functions mapping one bit to another, for example. Despite this apparent infinity, there exist universal sets of quantum gates, composed of a few specific gate types, that are capable of implementing any quantum computation. As a result, we can construct complex quantum circuits using only a limited number of gates from these universal sets to execute non-trivial quantum computations. This explains why the sets of common quantum gates typically include only a select few types.

In real-world quantum computing, we often focus on approximating desired computations to a controllable degree of accuracy. Hence, we define a universal gate set as follows:

Definition A *universal gate set* is a set of quantum gates that can approximate any unitary transformation on any number of qubits to an arbitrary degree of accuracy.

In simpler terms, a universal gate set allows us to approximate any desired quantum operation with a sequence of gates from the set, up to a predetermined error bound.

Universality is crucial in quantum computing because it allows us to solve any computational problem by using a fixed set of quantum gates, just as we can solve any computational problem in classical computing using a fixed set of classical gates.

So, what are the requirements for a universal set of quantum gates? At a minimum, they must allow us to reach any point on the Bloch sphere (including those with complex amplitudes), and they must be able to create both superposition states and entanglement states.

Below are a few examples of universal gate sets. Importantly, the choice of gate set depends on specific factors such as the quantum computing architecture, physical implementation, and system error rates. Although a universal gate set can approximate any quantum computation, it might not always provide the most efficient solution for certain tasks. Each gate set presents its own set of advantages and challenges. As such, research is ongoing to identify the optimal gate sets for various quantum computing platforms and applications.

7.2.1 CNOT + Arbitrary Single-Qubit Rotations

The combination of the CNOT gate and single-qubit unitary rotation gates accomplishes the requirements for a universal set. This gate set is universal because it can generate any single-qubit unitary transformation and create entanglement between qubits. (For a rigorous proof, see Refs. [18, 7].)

Single-qubit unitary rotations allow us to reach any point on the Bloch sphere, corresponding to any state the qubit can be in. Typically, two non-commuting rotation gates (representing rotations around different axes of the Bloch sphere) are sufficient to generate arbitrary rotations. A common choice involves the rotation gates around the x and z axes, denoted as R_x and R_z respectively, as defined in § 5.2.5.

In addition to the CNOT gate, frequently used gates in this universal set include X , Z , and H . While these are essential, other single-qubit gates may also be required for specific applications or implementations. The versatility and power of this universal set lie in its ability to perform both fundamental quantum operations and more complex algorithms across various quantum computing platforms.

7.2.2 Clifford Group + T

The Clifford Group, T gate set is a commonly utilized universal gate set, especially important in the context of fault-tolerant quantum computing, where fault-tolerance refers to the ability of a system to correct errors that occur during quantum computations. The Clifford group here encompasses the Pauli gates (X , Y , Z), the Hadamard gate (H), the phase gate (S), and the CNOT gate. These gates can be implemented efficiently and with fault-tolerance in many quantum error-correcting codes.

The T gate, representing a $\pi/4$ rotation around the z -axis of the Bloch sphere, is not a part of the Clifford group. It is specifically included in this set to efficiently approximate all single-qubit rotations, thus ensuring universality. In comparison to the Clifford gates, the T gate is more challenging to implement fault-tolerantly, thereby making it a major contributor to the computational cost of a quantum algorithm in fault-tolerant settings.

The T gate is harder to implement fault-tolerantly because it requires precise control over the qubit state, which can be challenging due to issues such as quantum decoherence and gate errors. This distinction has led to the concept of the T -count, which measures the number of T gates used in a quantum algorithm, serving as an indicator of the algorithm's computational expense in a fault-tolerant quantum computing paradigm. The universality of this gate set, coupled with its fault-tolerant characteristics, underpins its widespread usage.

7.2.3 Toffoli + H

The gate set comprised of the Toffoli gate (also known as the CCNOT gate) and Hadamard (H) gate forms a universal gate set for quantum computation. The Toffoli gate is a three-qubit gate, and is the most elementary quantum gate that is capable of executing deterministic universal classical computation. See § 7.5.1 for more details.

This gate set can produce any unitary transformation on any number of qubits, making it universal for quantum computation. It allows for the effective utilization of all inherent quantum mechanical properties such as superposition, entanglement, and quantum interference, providing the computational advantage sought in quantum computing.

The main practical challenge with this gate set lies in the implementation of the Toffoli gate. Because the Toffoli gate is a three-qubit gate, it often requires more complex quantum operations and higher error rates compared to single- and two-qubit gates. To implement a Toffoli gate, one might need to use a combination of simpler gates, which increases the gate count and thereby the possibility of errors.

7.3 Boolean Representation of Quantum Gates

In this section, we introduce a powerful tool for analyzing quantum gates and circuits: the Boolean representation of quantum gates and circuits. At its core, this representation offers an efficient method for us to examine a quantum circuit for each of the basis states separately. The results for each basis state are then combined to form a complete picture of the circuit's behavior for a general quantum state, leveraging the linearity property of quantum transformations. We also refer to this approach the logic operation approach.

7.3.1 Properties of XOR Operations

Consider x as a Boolean variable where $x \in \{0, 1\}$, and let $\bar{x} \equiv 1 - x$. When $x = 0$, $|x\rangle$ represents $|0\rangle$, and $|\bar{x}\rangle$ represents $|1\rangle$. When $x = 1$, $|x\rangle$ represents $|1\rangle$, and $|\bar{x}\rangle$ represents $|0\rangle$.

The XOR operator, denoted as \oplus , corresponds to a bitwise modulo 2 addition:

$$x \oplus y \equiv (x + y) \bmod 2. \quad (7.27)$$

Because $2x = 2y = 0 \bmod 2$, we also have

$$x \oplus y = (x - y) \bmod 2 = (y - x) \bmod 2. \quad (7.28)$$

The XOR operation results in ‘1’ when the number of ‘1’s in the operands is odd; otherwise, it returns ‘0’. (For example, $1 \oplus 1 \oplus 0 \oplus 1 = 1$, and $1 \oplus 1 \oplus 1 \oplus 1 = 0$.)

The following lists some key properties of XOR operations:

$$x \oplus 0 = x, \quad (7.29a)$$

$$x \oplus 1 = \bar{x}, \quad (7.29b)$$

$$\bar{x} \oplus 1 = x, \quad (7.29c)$$

$$x \oplus x = 0, \quad (7.29d)$$

$$x \oplus \bar{x} = 1, \quad (7.29e)$$

$$\overline{x \oplus y} = \bar{x} \oplus y = x \oplus \bar{y} = 1 \oplus x \oplus y, \quad (7.29f)$$

$$(-1)^{x \oplus y} = (-1)^{\bar{x} \oplus \bar{y}} = (-1)^{x+y}, \quad (7.29g)$$

$$(-1)^{x \oplus \bar{y}} = (-1)^{\bar{x} \oplus y} = (-1)^{1+x+y}, \quad (7.29h)$$

$$x \oplus y = x + y - 2xy. \quad (7.29i)$$

Exercise 7.8 Verify the last three relations in Eq. 7.29 for all combinations of $x = 0, 1$ and $y = 0, 1$.

Exercise 7.9 Simplify $\overline{x \oplus \bar{y} \oplus 1}$.

Exercise 7.10 Simplify $(-1)^{\overline{1 \oplus x \oplus \bar{y}}}$.

7.3.2 Boolean Representation of Common Quantum Gates

The Boolean representation of a quantum gate can be thought of as a formula that succinctly characterizes the transformation applied by the gate to all possible basis states. This is similar to the use of Boolean algebra and truth tables in classical computing, with significant differences which we will explain in § 7.3.3.

The Boolean representation of some common quantum gates is given in Table 7.2.

Gate	Symbol	Boolean Representation
X (NOT)	$ x\rangle \xrightarrow{X} \bar{x}\rangle$	$ x\rangle \mapsto \bar{x}\rangle$
CNOT	$\begin{array}{c} x\rangle \xrightarrow{\bullet} x\rangle \\ y\rangle \xrightarrow{\oplus} x \oplus y\rangle \end{array}$	$ xy\rangle \mapsto x(x \oplus y)\rangle$
Z	$ x\rangle \xrightarrow{Z} (-1)^x x\rangle$	$ x\rangle \mapsto (-1)^x x\rangle$
CZ	$\begin{array}{c} x\rangle \xrightarrow{\bullet} x\rangle \\ y\rangle \xrightarrow{Z} (-1)^{xy} y\rangle \end{array}$	$ xy\rangle \mapsto (-1)^{xy} xy\rangle$
H	$ x\rangle \xrightarrow{H} \frac{1}{\sqrt{2}} (0\rangle + (-1)^x 1\rangle)$	$ x\rangle \mapsto \frac{1}{\sqrt{2}} (0\rangle + (-1)^x 1\rangle)$
Y	$ x\rangle \xrightarrow{Y} i(-1)^x \bar{x}\rangle$	$ x\rangle \mapsto i(-1)^x \bar{x}\rangle$
$P(\phi)$	$ x\rangle \xrightarrow{e^{i\phi}} e^{ix\phi} x\rangle$	$ x\rangle \mapsto e^{ix\phi} x\rangle$
U	$ x\rangle \xrightarrow{U} u_{xx} x\rangle + u_{\bar{x}x} \bar{x}\rangle$	$ x\rangle \mapsto u_{xx} x\rangle + u_{\bar{x}x} \bar{x}\rangle$
CU	$\begin{array}{c} x\rangle \xrightarrow{\bullet} x\rangle \\ y\rangle \xrightarrow{U} U^x y\rangle, \text{ or} \\ u_{yy}^x y\rangle + xu_{\bar{y}y} \bar{y}\rangle \end{array}$	$ xy\rangle \mapsto x\rangle U^x y\rangle$ $ xy\rangle \mapsto u_{yy}^x xy\rangle + xu_{\bar{y}y} \bar{xy}\rangle$
Toffoli (CCNOT)	$\begin{array}{c} x\rangle \xrightarrow{\bullet} x\rangle \\ y\rangle \xrightarrow{\bullet} y\rangle \\ z\rangle \xrightarrow{\oplus} xy \oplus z\rangle \end{array}$	$ xyz\rangle \mapsto xy(xy \oplus z)\rangle$
SWAP	$\begin{array}{c} x\rangle \xrightarrow{\bullet} y\rangle \\ y\rangle \xrightarrow{\times} x\rangle \end{array}$	$ xy\rangle \mapsto yx\rangle$
Fredkin (CSWAP)	$\begin{array}{c} x\rangle \xrightarrow{\bullet} x\rangle \\ y\rangle \xrightarrow{\times} \bar{xy} \oplus xz\rangle \\ z\rangle \xrightarrow{\times} xy \oplus \bar{x}z\rangle \end{array}$	$ xyz\rangle \mapsto x(\bar{xy} \oplus xz)(xy \oplus \bar{x}z)\rangle$

Table 7.2: Boolean Representation of Common Qubit Gates

A note on the notation in Table 7.2. In the case of the CNOT gate, the expression $|x(x \oplus y)\rangle$ should be interpreted as $|x\rangle \otimes |(x \oplus y)\rangle$. This might appear unusual due to the omission of the tensor product operator for brevity. The juxtaposition of x and $(x \oplus y)$ within the ket notation does not imply multiplication. In this context, $|x(x \oplus y)\rangle$ represents states such as $|00\rangle$, $|11\rangle$, etc., with the first and second qubit values being x and $x \oplus y$, respectively.

In the Toffoli gate representation, the output of the third qubit is $|xy \oplus z\rangle$ in $|xyz\rangle \mapsto |xy(xy \oplus z)\rangle$. Here, the operation xy within the “($)$ ” denotes standard multiplication, not a tensor product. This is because $|xy \oplus z\rangle$ in this case

| represents a single qubit state.

Consider the Z gate as an example. Its Boolean representation is $|x\rangle \mapsto (-1)^x |x\rangle$. This formula means that for $x = 0$, we have $|0\rangle \mapsto |0\rangle$; and for $x = 1$, $|1\rangle \mapsto -|1\rangle$. Given the superposition state $\alpha|0\rangle + \beta|1\rangle$, we can apply the Z gate by leveraging the linearity of quantum transformations, which allows us to handle each term separately. Therefore, we obtain $\alpha|0\rangle + \beta|1\rangle \mapsto \alpha|0\rangle - \beta|1\rangle$, accurately reflecting the transformation function of the Z gate.

Exercise 7.11 Demonstrate that the Boolean representation of gate CZ is equal to that of CZ'. Here, CZ' represents the CZ gate with qubit 2 serving as the control.

Exercise 7.12 Formulate the Boolean representation of the parallel gate $I \otimes X$ and compare with the CNOT gate.

Exercise 7.13 Formulate the Boolean representation of the parallel gate $I \otimes Z$ and compare with the CZ gate.

7.3.3 Understanding Boolean Representation

While Boolean representation is a powerful technique for analyzing quantum circuits, it can sometimes lead to confusion during interpretation. To elucidate some key aspects and avoid potential pitfalls, we will examine specific examples next.

1 Forming Superposition

Boolean representation informs us the transformation of each basis state by a quantum gate or circuit. The results for each basis state are then combined to form a complete picture of the circuit's behavior for a general quantum state. Note that when multiple qubits are involved, the basis states are tensor product states.

■ Example 7.1 Let's use CNOT as an example to illustrate how the above process works. The symbol and formula for CNOT in Table 7.2 capture the transformation of CNOT on all combinations of the basis states as follows:

$$\begin{array}{llll} |00\rangle \mapsto |00\rangle & |01\rangle \mapsto |01\rangle & |10\rangle \mapsto |11\rangle & |11\rangle \mapsto |10\rangle \\ \begin{array}{c} |0\rangle \xrightarrow{\text{---}} |0\rangle \\ |0\rangle \xrightarrow{\oplus} |0\rangle \end{array} & \begin{array}{c} |0\rangle \xrightarrow{\text{---}} |0\rangle \\ |1\rangle \xrightarrow{\oplus} |1\rangle \end{array} & \begin{array}{c} |1\rangle \xrightarrow{\text{---}} |1\rangle \\ |0\rangle \xrightarrow{\oplus} |1\rangle \end{array} & \begin{array}{c} |1\rangle \xrightarrow{\text{---}} |1\rangle \\ |1\rangle \xrightarrow{\oplus} |0\rangle \end{array} \end{array}$$

Figure 7.4: Expansion of the Boolean Representation of CNOT

Combining these cases in a superposition, we obtain the general transformation of the CNOT gate:

$$\alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle \mapsto \alpha|00\rangle + \beta|01\rangle + \gamma|11\rangle + \delta|10\rangle.$$

2 $|x\rangle$ vs. $|\psi\rangle$

If a qubit has an output $|x\rangle$, we can treat $|x\rangle$ as the general quantum state $|\psi\rangle$, provided that qubit is disentangled from the rest of the system.

Care should be taken when dealing with multi-qubit systems. If $|x\rangle$ is entangled with another qubit, then we cannot directly substitute $|x\rangle$ with the general state $|\psi\rangle$ as $|x\rangle$ must be treated as a part of the multi-qubit basis states. Let's illustrate this caveat with the following example.

Example 7.2 A common question for the Boolean representation of CNOT gate is: If $|x\rangle \equiv |\psi\rangle$, doesn't the following diagram imply that we can duplicate an arbitrary state $|\psi\rangle$, thereby violating the no-cloning theorem?



The answer is NO. The replacement of $|x\rangle$ with $|\psi\rangle$ is invalid here since the two qubits are entangled. Thus, we need to instantiate the left part of the diagram twice, once with $x = 0$ and again with $x = 1$, and then perform a linear combination. Therefore, the diagram should be corrected as follows:

$$\begin{array}{ccc} |x\rangle & \xrightarrow{\quad\quad\quad} & \alpha |0\rangle + \beta |1\rangle \\ |0\rangle & \oplus & |0\rangle \end{array} \quad \Rightarrow \quad \left. \begin{array}{c} |x\rangle \\ |0\rangle \end{array} \right\} \alpha |00\rangle + \beta |11\rangle$$

■

3 * Identifying Product States

In the context of Boolean representation, where qubits are denoted as $|x\rangle$, $|y\rangle$, etc., we need to be cautious in discerning product states. A quantum state, even if it seems to present as a product state, may not be a true product state, as shown in the following example. The crux lies in understanding that only if a quantum state consistently preserves its form as a product state across all combinations of x and y values, inclusive of their signs, can it be treated as a genuine product state. In the absence of this uniformity across all basis states, the seeming product state is not a true one.

Example 7.3 This example comes from the discussion of quantum teleportation in § 10.3.3. At some stage in the process, the three-qubit system is in the state:

$$|\Psi_1\rangle = \frac{1}{\sqrt{2}}(|xx0\rangle + |x\bar{x}1\rangle). \quad (7.30)$$

At a first glance, you might think that the first qubit is separated from the system, because it seems that $|\Psi_1\rangle$ can be written as a product state $|x\rangle \otimes \frac{1}{\sqrt{2}}(|x0\rangle + |\bar{x}1\rangle)$.

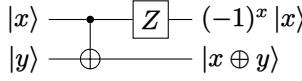
But this conclusion is wrong. Because of the presence of \bar{x} , the seemingly product state does not keep the same form for $x = 0$ and 1 .

To interpret Eq. 7.30 correctly, we need to run it twice, once for with $|x\rangle = |0\rangle$ and the other with $|x\rangle = |1\rangle$, and then do a linear combination of the two resulting equations with weights α and β . Thus, Eq. 7.30 represents

$$|\Psi_1\rangle = \frac{1}{\sqrt{2}} (\alpha(|000\rangle + |011\rangle) + \beta(|110\rangle + |101\rangle)), \quad (7.31)$$

which cannot be decomposed as the tensor product of $\alpha|0\rangle + \beta|1\rangle$ with another state. ■

Exercise 7.14 What is the output state from the following circuit when the input state is $|x\rangle = \alpha|0\rangle + \beta|1\rangle$ and $|y\rangle = \gamma|0\rangle + \delta|1\rangle$?



4 * Working with Coefficients

The following example illustrates how coefficients (of $|x\rangle$, etc.) work in Boolean representation.

■ **Example 7.4** Let's examine the CU gate in the last row of Table 7.2.

$$\begin{aligned} |xy\rangle &\mapsto u_{yy}^x |xy\rangle + xu_{\bar{y}\bar{y}} |\bar{x}\bar{y}\rangle \\ &= \begin{cases} |0y\rangle & \text{for } x = 0 \\ u_{yy} |1y\rangle + u_{\bar{y}\bar{y}} |1\bar{y}\rangle & \text{for } x = 1 \end{cases} \\ &= \begin{cases} |00\rangle & \text{for } x = 0, y = 0 \\ |01\rangle & \text{for } x = 0, y = 1 \\ u_{00} |10\rangle + u_{10} |11\rangle & \text{for } x = 1, y = 0 \\ u_{11} |11\rangle + u_{01} |10\rangle & \text{for } x = 1, y = 1. \end{cases} \end{aligned}$$

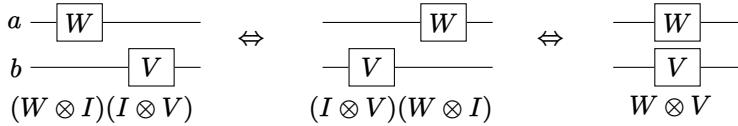
Exercise 7.15 Verify the Boolean representation of other gates in Table 7.2 as we have done for CU. ■

7.4 Equivalent Gate Sequences

In this section, we delve into a fundamental aspect of quantum circuits: gate sequences, and more importantly, their equivalence. The sequence of gates, much akin to the building blocks of a structure, forms the simplest embodiment of a quantum circuit. A fascinating attribute of these sequences is their ability to combine into a single gate equivalent, paving the path for simplification and optimization of quantum circuits. This becomes crucial in practical scenarios when the circuits need to be executed on actual quantum hardware where resources are limited and hence, optimization is key. Moreover, the understanding of equivalent gate sequences forms the bedrock for the analysis of more intricate quantum circuits, for instance, quantum gate teleportation. Mastering the techniques and skills required to dissect and interpret these sequences is a pivotal part of one's journey in quantum computing. As we proceed, we'll learn more about these sequences, their inherent properties, and the tools needed to comprehend their nuances.

7.4.1 Order Independence in Separate Qubit Operations

The figure below shows a sequence often implied without proof. Two linear operations, W and V , pertaining to two qubits a and b , respectively, are order independent. Here, W and V can represent either quantum gates or measurements.



Mathematically, applying W to qubit a without affecting qubit b is equivalent to applying $W \otimes I$ to the joint state of the two qubits. Similarly, applying V to qubit b only is equivalent to $I \otimes V$. In this case, W and V commute regardless of their commutation properties within their respective systems. This is because the operations act on different Hilbert spaces and do not interfere with each other. In other words,

$$(W \otimes I)(I \otimes V) = (I \otimes V)(W \otimes I), \text{ or } [W \otimes I, I \otimes V] = 0. \quad (7.32)$$

Because of this property, we often express the whole operation as $W \otimes V$ or $W_a V_b$.

Proof. Since

$$(W \otimes I)(I \otimes V) = WI \otimes IV = W \otimes V, \quad (7.33a)$$

$$(I \otimes V)(W \otimes I) = IW \otimes VI = W \otimes V, \quad (7.33b)$$

we have

$$[W \otimes I, I \otimes V] = (W \otimes I)(I \otimes V) - (I \otimes V)(W \otimes I) \quad (7.34a)$$

$$= W \otimes V - W \otimes V \quad (7.34b)$$

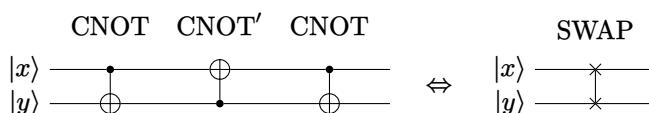
$$= 0. \quad (7.34c)$$

□

7.4.2 Gate Sequence: A First Example

Gate equivalence can be analyzed using three approaches: matrix product, operator product, and logic operation. Mastering these approaches can greatly aid in optimizing quantum circuits and algorithms. The choice of approach, whether more intuitive or effective, hinges on the nature of the gate sequence.

We will illustrate these approaches with a prominent example of gate equivalence, depicted in the diagram below. When sequentially applied, the trio of CNOT gates (with the central one flipped) effectively emulates the operation of a SWAP gate, which interchanges the states of two qubits.



1 Matrix Product Approach

The matrix product representing the gate sequence:

$$\text{CNOT} \cdot \text{CNOT}' \cdot \text{CNOT} \quad (7.35a)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (7.35b)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (7.35c)$$

We recognize that the final product matrix corresponds to the SWAP gate.

2 Operator Product Approach

Recall from §§ 7.1.1 and 7.1.2 that:

$$\text{CNOT} = |00\rangle\langle 00| + |01\rangle\langle 01| + |11\rangle\langle 10| + |10\rangle\langle 11|, \quad (7.36a)$$

$$\text{CNOT}' = |00\rangle\langle 00| + |11\rangle\langle 01| + |10\rangle\langle 10| + |01\rangle\langle 11|. \quad (7.36b)$$

The operator product representing the gate sequence:

$$\text{CNOT} \cdot \text{CNOT}' \cdot \text{CNOT} \quad (7.37a)$$

$$= (|00\rangle\langle 00| + |01\rangle\langle 01| + |11\rangle\langle 10| + |10\rangle\langle 11|) \quad (7.37b)$$

$$(|00\rangle\langle 00| + |11\rangle\langle 01| + |10\rangle\langle 10| + |01\rangle\langle 11|) \quad (7.37c)$$

$$(|00\rangle\langle 00| + |01\rangle\langle 01| + |11\rangle\langle 10| + |10\rangle\langle 11|) \quad (7.37d)$$

$$= (|00\rangle\langle 00| + |01\rangle\langle 01| + |11\rangle\langle 10| + |10\rangle\langle 11|) \quad (7.37e)$$

$$(|00\rangle\langle 00| + |11\rangle\langle 01| + |10\rangle\langle 11| + |01\rangle\langle 10|) \quad (7.37f)$$

$$= |00\rangle\langle 00| + |01\rangle\langle 10| + |10\rangle\langle 01| + |11\rangle\langle 11| \quad (7.37g)$$

$$= \text{SWAP}. \quad (7.37h)$$

3 Logic Operation Approach

Using the Boolean representation tool developed in § 7.3 we can derive the same equivalence as follows.

Recall from § 7.3 that:

$$\text{CNOT} : |x, y\rangle \mapsto |x, x \oplus y\rangle, \quad (7.38a)$$

$$\text{CNOT}' : |x, y\rangle \mapsto |x \oplus y, y\rangle, \quad (7.38b)$$

where $x, y \in \{0, 1\}$.

Now we go through the three quantum gates in the sequence:

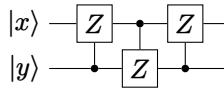
$$|x, y\rangle \xrightarrow{\text{first CNOT}} |x, x \oplus y\rangle \quad (7.39a)$$

$$\xrightarrow{\text{mid CNOT}'} |x \oplus (x \oplus y), x \oplus y\rangle = |y, x \oplus y\rangle \quad (7.39b)$$

$$\xrightarrow{\text{second CNOT}} |y, (x \oplus y) \oplus y\rangle = |y, x\rangle. \quad (7.39c)$$

We can see $|x, y\rangle$ maps to $|y, x\rangle$, which represents the SWAP operation.

Exercise 7.16 Replace CNOT with CZ in the above analysis, and investigate the equivalence of the following gate sequence. Identify if this sequence emulates any well-known gate operation, similar to how the previous sequence performed a SWAP operation:



7.4.3 Sequences Involving X, Z, and CNOT

The ordered gate sequences in this subsection, involving X , Z , and CNOT, are essential for understanding CNOT gate transportation discussed in § 10.5.2.

Seq.	Gate Sequence	Equivalent Sequence	Proof
1	$ x\rangle \xrightarrow{\text{---}} \bullet \quad y\rangle \xrightarrow[X]{\oplus} \bullet \quad \left\{ \right. y\rangle \left. \right\} \Psi_1\rangle$	$ x\rangle \xrightarrow{\text{---}} \bullet \quad y\rangle \xrightarrow[\oplus]{X} \bullet \quad \left\{ \right. y\rangle \left. \right\} \Psi_2\rangle$	Eq. 7.40
2	$ x\rangle \xrightarrow[X]{\bullet} \bullet \quad y\rangle \xrightarrow[\oplus]{\text{---}} \bullet \quad \left\{ \right. y\rangle \left. \right\} \Psi_1\rangle$	$ x\rangle \xrightarrow{\text{---}} \bullet \quad y\rangle \xrightarrow[\oplus]{X} \bullet \quad \left\{ \right. x\rangle \left. \right\} \Psi_2\rangle$	Eq. 7.41
3	$ x\rangle \xrightarrow{\text{---}} \bullet \quad y\rangle \xrightarrow[Z]{\oplus} \bullet \quad \left\{ \right. y\rangle \left. \right\} \Psi_1\rangle$	$ x\rangle \xrightarrow{\text{---}} \bullet \quad y\rangle \xrightarrow[\oplus]{Z} \bullet \quad \left\{ \right. x\rangle \left. \right\} \Psi_2\rangle$	Eq. 7.42
4	$ x\rangle \xrightarrow[Z]{\bullet} \bullet \quad y\rangle \xrightarrow[\oplus]{\text{---}} \bullet \quad \left\{ \right. y\rangle \left. \right\} \Psi_1\rangle$	$ x\rangle \xrightarrow{\text{---}} \bullet \quad y\rangle \xrightarrow[\oplus]{Z} \bullet \quad \left\{ \right. x\rangle \left. \right\} \Psi_2\rangle$	Eq. 7.43

Table 7.3: Gate Sequence Equivalence Involving X , Z and CNOT

The following equations provide proofs for the gate sequence equivalences referenced in Table 7.3. Here, C denotes the operator of the CNOT gate. We have used the logic operation approach for these proofs. Of course, we can also prove these equivalences using matrix multiplication. Please reference § 7.3 for the properties of the XOR (\oplus) operation and Boolean representations of quantum gates.

Gate Sequence 1

$$|\Psi_1\rangle = CX_2|x\rangle|y\rangle = C|x\rangle|\bar{y}\rangle = |x\rangle|x \oplus \bar{y}\rangle, \quad (7.40a)$$

$$|\Psi_2\rangle = X_2C|x\rangle|y\rangle = X_2|x\rangle|x \oplus y\rangle = |x\rangle|\overline{x \oplus y}\rangle. \quad (7.40b)$$

Here we have used $\overline{x \oplus y} = x \oplus \bar{y} = \bar{x} \oplus y = 1 \oplus x \oplus y$.

Given that $|\Psi_1\rangle$ and $|\Psi_2\rangle$ equal to $|x\rangle|\bar{x} \oplus y\rangle$, these two gate sequences also equivalent to CNOT, the inverted CNOT where the NOT gate is activated when the control qubit is $|0\rangle$, as opposed to $|1\rangle$. See also Problem 7.1. This equivalence can also be derived through matrix multiplication.

Gate Sequence 2

$$|\Psi_1\rangle = CX_1|x\rangle|y\rangle = C|\bar{x}\rangle|y\rangle = |\bar{x}\rangle|\bar{x} \oplus y\rangle, \quad (7.41a)$$

$$|\Psi_2\rangle = X_2X_1C|x\rangle|y\rangle = X_2X_1|x\rangle|x \oplus y\rangle = |\bar{x}\rangle|\overline{x \oplus y}\rangle. \quad (7.41b)$$

Here we have used $\overline{x \oplus y} = \bar{x} \oplus y = 1 \oplus x \oplus y$.

Gate Sequence 3

$$|\Psi_1\rangle = CZ_2|x\rangle|y\rangle = (-1)^yC|x\rangle|y\rangle = (-1)^y|x\rangle|x \oplus y\rangle, \quad (7.42a)$$

$$|\Psi_2\rangle = Z_2Z_1C|x\rangle|y\rangle = Z_2Z_1|x\rangle|x \oplus y\rangle = (-1)^{x+x \oplus y}|x\rangle|x \oplus y\rangle. \quad (7.42b)$$

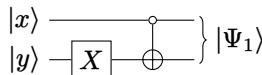
Here we have used $(-1)^{x+x \oplus y} = (-1)^{x+x+y} = (-1)^{2x+y} = (-1)^y$.

Gate Sequence 4

$$|\Psi_1\rangle = CZ_1|x\rangle|y\rangle = (-1)^xC|x\rangle|y\rangle = (-1)^x|x\rangle|x \oplus y\rangle, \quad (7.43a)$$

$$|\Psi_2\rangle = Z_1C|x\rangle|y\rangle = Z_1|x\rangle|x \oplus y\rangle = (-1)^x|x\rangle|x \oplus y\rangle. \quad (7.43b)$$

Exercise 7.17 * The diagram below depicts a gate sequence similar to the sequence 1 in Table 7.3, except CNOT is replaced with $\overline{\text{CNOT}}$. The $\overline{\text{CNOT}}$ gate is an inverted CNOT gate: it applies an X gate to the target qubit when the control qubit is in state $|0\rangle$, instead of $|1\rangle$. Find an equivalent sequence to this sequence, similar to the ones in Table 7.3.

**7.4.4 Sequences Involving H**

The gate sequences detailed in Table 7.4 incorporate the Hadamard gate H . For these sequences, analysis via matrix multiplication tends to be the most straightforward approach.

Seq.	Gate Sequence	Equivalent Sequence	Proof
5			Eq. 7.44
6			Eq. 7.45
7			Eq. 7.46
8			Eq. 7.47

Table 7.4: Gate Sequence Equivalence Involving H

The subsequent equations furnish proofs for the gate sequence equivalences cited in Table 7.4. For each case, we use U to denote the unitary matrix representing the gate sequence.

Gate Sequence 5

$$U = (I \otimes H) \text{CNOT}(I \otimes H) \quad (7.44\text{a})$$

$$= \begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & X \end{bmatrix} \begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix} \quad (7.44\text{b})$$

$$= \begin{bmatrix} H^2 & 0 \\ 0 & HXH \end{bmatrix} \quad (7.44\text{c})$$

$$= \begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix} \quad (7.44\text{d})$$

$$= \text{CZ}. \quad (7.44\text{e})$$

Here we have used $H^2 = I$ and $HXH = Z$.

Gate Sequence 6

$$U = (H \otimes I) \text{CZ}(H \otimes I) \quad (7.45\text{a})$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \quad (7.45\text{b})$$

$$= \frac{1}{2} \begin{bmatrix} I+Z & I-Z \\ I-Z & I+Z \end{bmatrix} \quad (7.45\text{c})$$

$$= \text{CNOT}' . \quad (7.45\text{d})$$

Here we have used Eq. 7.8 for the matrix of CNOT' .

Gate Sequence 7

$$U = (H \otimes H) \text{CNOT}(H \otimes H) \quad (7.46\text{a})$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} H & H \\ H & -H \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & X \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} H & H \\ H & -H \end{bmatrix} \quad (7.46\text{b})$$

$$= \frac{1}{2} \begin{bmatrix} I+Z & I-Z \\ I-Z & I+Z \end{bmatrix} \quad (7.46\text{c})$$

$$= \text{CNOT}' . \quad (7.46\text{d})$$

Here we have used Eq. 7.8 for the matrix of CNOT' .

You may have observed that gate sequences 6 and 7 are equivalent, both corresponding to CNOT' . Interestingly, another method to deduce sequence 7 is by combining sequences 5 and 6, and applying the identity $H^2 = I$. We leave this derivation to the reader.

Gate Sequence 8

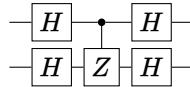
As it turns out, gate sequence 8, even though appearing very similar to sequence 6, does not correspond to a single equivalent gate within our current repertoire. Its equivalent form shown in Table 7.4 can be derived from sequence 7. Its operator is given by:

$$U = (H \otimes I) \text{CNOT}(H \otimes I) \quad (7.47a)$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & X \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \quad (7.47b)$$

$$= \frac{1}{2} \begin{bmatrix} I+X & I-X \\ I-X & I+X \end{bmatrix}. \quad (7.47c)$$

Exercise 7.18 You're given the following gate sequence, which is similar to sequence 7 in Table 7.4 except the CNOT is replaced by CZ.



Your tasks are two-fold:

(a) Simplify this sequence using the existing equivalent sequences found in Table 7.4. Provide at least two different ways to simplify this sequence.

(b) Verify your simplifications by performing matrix multiplication. The aim is to demonstrate that your simplified sequence is indeed equivalent to the original one.

7.4.5 *Deferred Measurement Principle

In this section, we will explore a type of gate sequence equivalence that involves measurements. This is enabled by the deferred measurement principle, which allows us to measure control qubits early in a quantum computation. As a result, quantum controlled gates can often be replaced with their classically controlled analogs. Fig. 7.5 provides an example of equivalent circuits that make use of this principle.

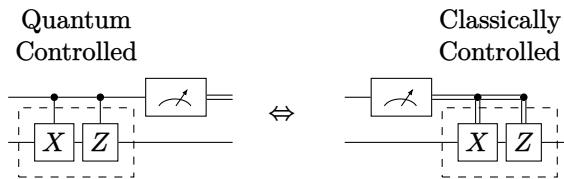
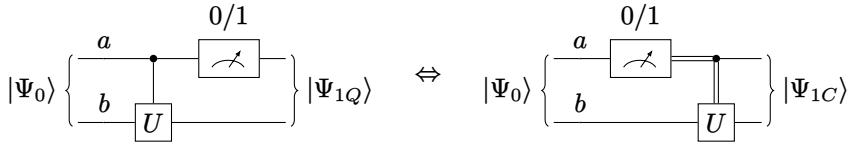


Figure 7.5: Example of Deferred Measurements

The Deferred Measurement Principle.

Measurements can be moved from an intermediate stage of a quantum circuit to the end. If the results of these measurements are used at any point in the circuit, classically controlled operations can be replaced by quantum controlled operations.

Mathematically, the deferred measurement principle signifies that the two circuits in Fig. 7.6 are equivalent, indicating that control (for a quantum Controlled- U gate) and measurement (whose outcome is utilized in the corresponding classically controlled U gate) commute.

**Figure 7.6:** Deferred Measurement Principle

Proof. * To prove the equivalence of the two circuits in Fig. 7.6, we need to show that their output states, i.e., $|\Psi_{1C}\rangle$ and $|\Psi_{1Q}\rangle$, are identical.

The measurement operator for both circuits is:

$$M_i = |i\rangle\langle i| \otimes I, \quad (7.48)$$

where i is the measurement outcome, 0 or 1.

The operator for a quantum controlled U gate, CU, is provided in Eq. 7.10; and cCU for a classically controlled U gate is in Eq. 7.20.

For the circuit on the left with a quantum controlled U gate:

$$|\Psi_{1Q}\rangle = \text{CU} M_i |\Psi_0\rangle \quad (7.49a)$$

$$= (|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes U)(|i\rangle\langle i| \otimes I) |\Psi_0\rangle \quad (7.49b)$$

$$= (|0\rangle\langle 0| |i\rangle\langle i|) \otimes (II) |\Psi_0\rangle + (|1\rangle\langle 1| |i\rangle\langle i|) \otimes (UI) |\Psi_0\rangle \quad (7.49c)$$

$$= \begin{cases} (|0\rangle\langle 0| \otimes I) |\Psi_0\rangle & \text{for } i = 0, \\ (|1\rangle\langle 1| \otimes U) |\Psi_0\rangle & \text{for } i = 1. \end{cases} \quad (7.49d)$$

For the circuit on the right with a classically controlled U gate:

$$|\Psi_{1C}\rangle = M_i \text{cCU} |\Psi_0\rangle \quad (7.50a)$$

$$= (|i\rangle\langle i| \otimes I)(I \otimes U^i) |\Psi_0\rangle \quad (7.50b)$$

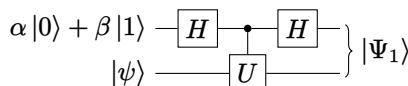
$$= (|i\rangle\langle i| I) \otimes (IU^i) |\Psi_0\rangle \quad (7.50c)$$

$$= (|i\rangle\langle i| \otimes U^i) |\Psi_0\rangle \quad (7.50d)$$

$$= \begin{cases} (|0\rangle\langle 0| \otimes I) |\Psi_0\rangle & \text{for } i = 0, \\ (|1\rangle\langle 1| \otimes U) |\Psi_0\rangle & \text{for } i = 1. \end{cases} \quad (7.50e)$$

Therefore, we conclude that $|\Psi_{1C}\rangle = |\Psi_{1Q}\rangle$. □

Exercise 7.19 * The following circuit is a basic building block in error correction codes.



(a) Demonstrate that the output state can be written as:

$$|\Psi_1\rangle = \frac{\alpha + \beta}{\sqrt{2}} (|0\rangle|\psi\rangle + |1\rangle|\psi\rangle) + \frac{\alpha - \beta}{\sqrt{2}} (|0\rangle U|\psi\rangle - |1\rangle U|\psi\rangle).$$

- (b) If the first qubit is measured and the outcome is $|0\rangle$ (or $|1\rangle$), what is the state of the second qubit upon measurement?

7.4.6 *Operator Sum Relations

In quantum circuits, gates are applied sequentially, corresponding to the multiplication of the respective matrices, not their addition. This distinction means that certain equivalences between gates, while mathematically accurate, cannot be translated directly into equivalent sequences of quantum gates.

A notable example of this is that some quantum gates can be expressed in terms of Pauli operators:

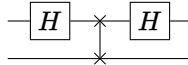
$$H = \frac{1}{\sqrt{2}}(X + Z), \quad (7.51a)$$

$$\text{CNOT} = \frac{1}{2}(I \otimes I + I \otimes X + Z \otimes I - Z \otimes X), \quad (7.51b)$$

$$\text{CZ} = \frac{1}{2}(I \otimes I + I \otimes Z + Z \otimes I - Z \otimes Z), \quad (7.51c)$$

$$\text{SWAP} = \frac{1}{2}(I \otimes I + X \otimes X + Y \otimes Y + Z \otimes Z). \quad (7.51d)$$

These relations are not directly translatable into quantum circuits in general as they are sums of operators, not compositions. However, They can still provide useful insights into the underlying operation of the gates. As an example, let's find an equivalent for the following gate sequence using the above equations:



$$(H \otimes I) \cdot \text{SWAP} \cdot (H \otimes I) \quad (7.52a)$$

$$= (H \otimes I) \frac{1}{2}(I \otimes I + X \otimes X + Y \otimes Y + Z \otimes Z)(H \otimes I) \quad (7.52b)$$

$$= \frac{1}{2}(HIH \otimes I + HXH \otimes X + HYH \otimes Y + HZH \otimes Z) \quad (7.52c)$$

$$= \frac{1}{2}(I \otimes I + Z \otimes X - Y \otimes Y + X \otimes Z) \quad (7.52d)$$

$$= \frac{1}{2} \begin{bmatrix} I + X & Z + iY \\ Z - iY & I - X \end{bmatrix} \quad (7.52e)$$

$$= \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \quad (7.52f)$$

Here we have used the following identities:

$$HXH = Z, \quad (7.53a)$$

$$HZH = X, \quad (7.53b)$$

$$HYH = -Y, \quad (7.53c)$$

and

$$(A \otimes B)(C \otimes D) = AC \otimes BD. \quad (7.54)$$

7.5 *Exploratory Topics

In this section, we delve into advanced topics that might not be essential for those new to quantum computing but can provide additional insights and depth for more advanced readers.

7.5.1 *Expressing Common Gates with the Toffoli Gate

In this subsection, we investigate the properties of the Toffoli (CCNOT) gate, as shown in Table 7.2. The Boolean representation of the Toffoli gate is given as

$$\text{Toffoli: } |xyz\rangle \mapsto |xy(xy \oplus z)\rangle. \quad (7.55)$$

The classical logic functions such as NOT, AND, and FANOUT can be expressed in terms of the Toffoli gate alone as demonstrated in Table 7.5. An OR gate can be represented using several stages of Toffoli gates following the relation $\text{OR}(x, y) = \text{NOT}(\text{AND}(\text{NOT}(x), \text{NOT}(y)))$. This demonstrates that we have a quantum equivalent for basic classical logic functions. An additional ancillary qubit, introduced in the Toffoli gate, makes this translation possible.

By establishing a connection with the universal set of classical logic gates—NOT, AND, and FANOUT—we can translate any classical logic circuit into a quantum circuit using only the Toffoli gate. This implies that a quantum computer can emulate the performance of a classical computer, at minimum, if hardware is available.

However, to utilize the full potential of quantum computing, we need gates that facilitate superposition and entanglement. The Toffoli gate can act as a CNOT gate, making it suitable for entanglement operations. When coupled with the Hadamard (H) gate, which allows for superposition, we obtain a universal set of quantum gates. This powerful combination, discussed in § 7.2.3, signifies that the Toffoli gate and Hadamard gate together can be utilized for any quantum computation.

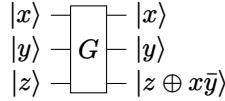
Gate	Symbol	Boolean Representation
NOT (X)	$ 1\rangle \xrightarrow{\bullet} 1\rangle$	
	$ 1\rangle \xrightarrow{\bullet} 1\rangle$	$ z\rangle \mapsto \bar{z}\rangle$
	$ z\rangle \xrightarrow{\oplus} \bar{z}\rangle$	
AND	$ x\rangle \xrightarrow{\bullet} x\rangle$	
	$ y\rangle \xrightarrow{\bullet} y\rangle$	$ xy\rangle \mapsto z = (xy)\rangle$
	$ 0\rangle \xrightarrow{\oplus} (xy)\rangle$	
FANOUT	$ 1\rangle \xrightarrow{\bullet} 1\rangle$	Only for classical bits represented by basis states; not for superposition states.
	$ y\rangle \xrightarrow{\bullet} y\rangle$	
	$ 0\rangle \xrightarrow{\oplus} y\rangle$	
CNOT	$ 1\rangle \xrightarrow{\bullet} 1\rangle$	
	$ y\rangle \xrightarrow{\bullet} y\rangle$	$ yz\rangle \mapsto y(y \oplus z)\rangle$
	$ z\rangle \xrightarrow{\oplus} (y \oplus z)\rangle$	

Table 7.5: Expressing Common Gates Using Toffoli Gate

Exercise 7.20 Express NOT, AND, and classical FANOUT using the Fredkin (CSWAP) gate exclusively, similar to the Toffoli gate. Evaluate if the Fredkin gate alone can construct a CNOT or H gate.

7.5.2 *Advanced Quantum Gate Examples

Suppose we need a quantum gate to perform for the following function:



Is this gate realizable as a quantum gate? In order to examine this, let's work out the mappings of the basis states:

$$\begin{aligned} |000\rangle &\mapsto |000\rangle, & |001\rangle &\mapsto |001\rangle, & |010\rangle &\mapsto |010\rangle, & |011\rangle &\mapsto |011\rangle, \\ |100\rangle &\mapsto |101\rangle, & |101\rangle &\mapsto |100\rangle, & |110\rangle &\mapsto |110\rangle, & |111\rangle &\mapsto |111\rangle. \end{aligned}$$

We notice the basis states $|100\rangle$ and $|101\rangle$ are changed, while the others remain unchanged. This mapping corresponds to the following matrix for the gate:

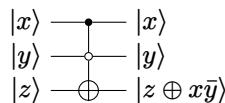
$$G = \begin{bmatrix} I_4 & 0 & 0 \\ 0 & U & 0 \\ 0 & 0 & I_2 \end{bmatrix},$$

where I_4 is the 4×4 identity matrix, I_2 is the 2×2 identity matrix, and U is a 2×2 matrix defined by:

$$U = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (7.56)$$

We recognize $U = X$. Apparently G is a unitary matrix. So it is realizable as a quantum gate.

What kind of gate is G ? The changed basis states, $|100\rangle \mapsto |101\rangle$ and $|101\rangle \mapsto |100\rangle$, imply that the third qubit z is flipped only when the first qubit $x = 1$ and the second qubit $y = 1$. From this we conclude that G can be named the C \bar{C} NOT gate, with a symbol:



This gate can be regarded as a variation of the original CCNOT (Toffoli) gate in Table 7.2.

Exercise 7.21 Given the matrix representation of a gate

$$U = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Find its corresponding Boolean representation.

7.6 Summary and Conclusions

Multi-Qubit Operations

The heart of this chapter revolved around multi-qubit gates, the operational bedrock upon which complex quantum circuits are constructed. These gates, acting on the tensor product of state spaces of the involved qubits, stand distinct from their single-qubit counterparts. One prominent aspect is their ability to generate and manipulate entangled states. While these gates possess the power to intricately intertwine qubits, they, like single-qubit gates, still represent unitary evolutions, ensuring conservation of total probability post-application.

Delving into the specifics, we embarked on a comprehensive study of familiar multi-qubit gates such as CNOT, CZ, and CU, highlighting their pivotal roles in quantum circuit design and algorithms. The structure and application of these gates were emphasized, setting the groundwork for intricate quantum computing tasks.

Universal Quantum Computing Framework

Drawing a parallel to classical computing, where complex operations evolve from simpler ones, the chapter transitioned to universal sets of qubit gates. The universality concept in quantum computing stems from the necessity to approximate any quantum operation using a restricted gate set. By introducing the reader to diverse universal gate sets and elaborating on their prerequisites, the chapter emphasized the importance of adaptability in the quantum realm, especially when harnessing limited gate types for non-trivial quantum computations.

Analytical Tools and Techniques

A central tenet of the chapter was the introduction of the Boolean representation of quantum gates and circuits. This methodology allows for efficient analysis of quantum circuits, considering each basis state independently, then combining them for a holistic understanding, thereby harnessing the linearity of quantum transformations. This technique not only aids in conceptual clarity but also proves instrumental in dissecting complex quantum circuits.

Venturing into gate sequences, we underlined the notion of equivalence in gate sequences. Recognizing sequences that combine to form a singular gate equivalent is essential for understanding complex quantum circuits, as well as for optimization, particularly when addressing real-world quantum hardware constraints.

Advanced Explorations

In this chapter, we also introduced some exploratory topics to provide a deeper understanding of quantum gates and their nuances. While these discussions might be seen as advanced, they aim to provide a foundational understanding for readers

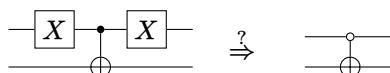
interested in further exploring the subject of quantum mechanics.

Upcoming Topics

In the forthcoming chapter, we journey into the realm of Bell states. These states, defined amidst the backdrop of the intriguing EPR paradox and Bell's theorem, stand as beacons elucidating quantum mechanics' complexities. As quintessential examples of quantum entanglement, Bell states not only accentuate quantum superposition and nonlocal correlations but also form the backbone of numerous quantum information processing tasks. This sets the stage for an enriching discourse on the EPR paradox and the myriad applications of quantum entanglement.

Problem Set 7

- 7.1** Work out the operator and matrix for a variant of the CNOT gate, denoted as $\bar{\text{CNOT}}$. In this gate, qubit 1 serves as the control qubit and qubit 2 as the target qubit. However, in contrast to the regular CNOT gate, an X (i.e., NOT) gate is applied to the target qubit, flipping its states 0 and 1, when the control qubit is in state $|0\rangle$ rather than $|1\rangle$.
- 7.2** Intuitively, you should be able to achieve the function of $\bar{\text{CNOT}}$ in Problem 7.1 by adding an X gate on the control qubit of a regular CNOT to flip the $|0\rangle$ and $|1\rangle$, and appending another X gate to restore the flip:



In terms of operators, this means

$$\bar{\text{CNOT}} = (X \otimes I) \cdot \text{CNOT} \cdot (X \otimes I).$$

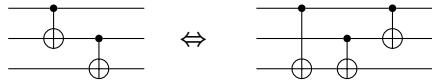
Prove or disprove the above equation using the $\bar{\text{CNOT}}$ operator and/or matrix you have obtained in Problem 7.1.

- 7.3** A CU gate applies a generic unitary gate U to the target qubit if the control qubit is $|1\rangle$. A $\bar{\text{CU}}$ gate applies U to the target qubit if the control qubit is $|0\rangle$. Work out the operator and matrix for $\bar{\text{CU}}$.
- 7.4** Derive the operators corresponding to the quantum gates defined by the following block matrices, where U is a 2×2 unitary matrix, and I the 2×2 identity matrix. For each gate, discuss its properties similarly to the approach used for CU in § 7.1.3. Analyze the patterns within each operator, formulate a description analogous to that of CU, and devise a corresponding circuit diagram.

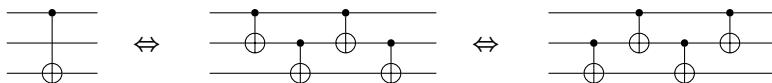
$$G_1 = \begin{bmatrix} 0 & U \\ I & 0 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & I \\ U & 0 \end{bmatrix}.$$

- 7.5** Derive the matrix of $R_{xy}(\theta)$, similar to Eq. 7.24.

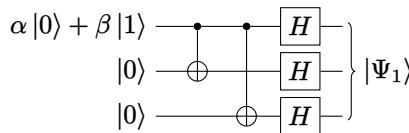
- 7.6 (CNOT Mirror)** The circuit equivalence shown below demonstrates that a pair of CNOT gates on three qubits—where the target qubit of the first CNOT is the control qubit of the second CNOT—can be rearranged. By reversing the sequence of these two CNOT gates and adding a CNOT at the beginning, the overall action remains unchanged. Prove this equivalence.



- 7.7 (Distributed CNOT)** In the distributed CNOT protocol, the CNOT operation is performed between two qubits without direct interaction by using an intermediate qubit and four CNOT gates, as illustrated by the equivalent circuits below. Prove this equivalence.



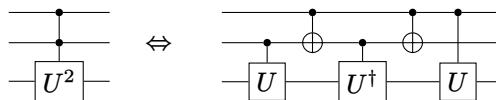
- 7.8** Examine the following circuit, which represents a basic encoder in error correction codes designed to rectify phase-inversion errors.

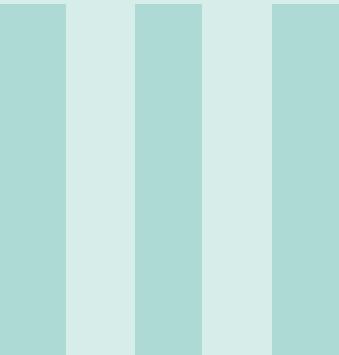


Demonstrate that the output state is:

$$|\Psi_1\rangle = \alpha |+++ \rangle + \beta |--- \rangle.$$

- 7.9** The following gate sequence is often used to decompose three-qubit gates into two-qubit gates. Prove its equivalence.





Quantum Entanglement

8	Bell States	201
9	Entanglement and Bell Inequalities	217
10	Key Applications of Entanglement	245

8. Bell States

Contents

8.1	Maximally Entangled States	202
8.1.1	Definitions	202
8.1.2	Maximum Entanglement	203
8.2	* Bell Basis	203
8.2.1	* Orthonormality and Completeness	204
8.2.2	* Two-Qubit States in Bell Basis	204
8.3	Bell State Creation	205
8.3.1	Basic Bell State Generator	205
8.3.2	Generating All Bell States from $ 00\rangle$	206
8.3.3	Photon Pair Creation with SPDC	207
8.3.4	* Boolean Representation	208
8.4	Bell Measurement	208
8.4.1	Joint Measurement	208
8.4.2	Basic Implementation: Bell Analyzer	209
8.4.3	Post-Measurement State	209
8.5	Bell State Conversion	209
8.5.1	Converting $ \beta_{00}\rangle$ to Other Bell States	210
8.5.2	Converting Other Bell States to $ \beta_{00}\rangle$	210
8.5.3	Additional Conversions Related to $ \beta_{11}\rangle$	211
8.6	* Deferred Proofs	211
8.7	* Generalization: GHZ States	212
8.8	Summary and Conclusions	213
	Problem Set 8	214

Bell states were first defined in the context of the famous “EPR paradox,” a thought experiment devised by Einstein, Podolsky, and Rosen in 1935 to argue that quantum mechanics was incomplete. It wasn’t until 1964 that John Bell formulated what is now known as Bell’s theorem, which showed that the kind of “hidden variables” suggested by EPR could not explain the predictions of quantum mechanics, and therefore quantum mechanics is a complete theory.

The Bell states were then used to test Bell's theorem experimentally, and the results of these tests have been overwhelmingly in favor of quantum mechanics. These tests, and the Bell states they rely on, are a major reason why physicists today accept quantum mechanics as a correct and complete description of the microscopic world.

Bell states are a set of four distinct entangled quantum states of two qubits. They are the simplest and most widely recognized examples of quantum entanglement — a remarkable quantum phenomenon wherein the properties of one qubit become intricately connected to the properties of another, irrespective of their spatial separation.

Bell states, being maximally entangled, serve to elucidate fundamental aspects of quantum mechanics such as quantum superposition and nonlocal correlations. Furthermore, Bell states play an indispensable role in a myriad of quantum computing and quantum information processing applications, including quantum teleportation, quantum cryptography, and quantum error correction.

In this chapter, we delve into a detailed exploration of Bell states, setting the stage for our discussion on the EPR paradox and applications of quantum entanglement in subsequent chapters. This also allows us to put into practice the concepts and techniques we have thus far studied concerning quantum gates.

8.1 Maximally Entangled States

8.1.1 Definitions

The four Bell states are:

$$|\Phi^+\rangle \equiv |\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \quad (8.1a)$$

$$|\Psi^+\rangle \equiv |\beta_{01}\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \quad (8.1b)$$

$$|\Phi^-\rangle \equiv |\beta_{10}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \quad (8.1c)$$

$$|\Psi^-\rangle \equiv |\beta_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \quad (8.1d)$$

These states are all non-product states, representing different types of correlations between the qubits.



In the context of quantum physics, the $|\Psi^-\rangle$ state is often referred to as a “spin singlet” while the other three states, “spin triplets”. The reason for this terminology is that, when we interpret the qubits as spin-1/2 particles, the $|\Psi^-\rangle$ state corresponds to a pair of particles with a total combined spin of 0. Conversely, the other three states represent a combined particle pair with a total spin of 1.

Exercise 8.1 Show that $\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$ is a product state.

Exercise 8.2 Show that $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is a non-product state.

8.1.2 Maximum Entanglement

Bell states are maximally entangled states of two qubits. Consequently, when we perform local measurements on the qubits in a Bell state, the measurement outcomes are perfectly correlated (or anti-correlated) due to their entanglement.

The measurement operators acting on the first qubit in the standard basis:

$$M_0 = |0\rangle\langle 0| \otimes I \quad \text{for measurement outcome } |0\rangle, \quad (8.2a)$$

$$M_1 = |1\rangle\langle 1| \otimes I \quad \text{for measurement outcome } |1\rangle. \quad (8.2b)$$

Let's take the $|\Phi^+\rangle$ Bell state as an example. If we measure the first qubit and find it in state $|0\rangle$, we have effectively applied the M_0 measurement operator:

$$M_0 |\Phi^+\rangle = (|0\rangle\langle 0| \otimes I) \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle). \quad (8.3)$$

This results in the post-measurement state $|00\rangle$:

$$\frac{1}{\sqrt{2}} (|00\rangle + |11\rangle) \xrightarrow{\text{measure 1st qubit to be } |0\rangle} |00\rangle. \quad (8.4)$$

As a result, we can be certain that the second qubit will also be in the state $|0\rangle$.

Similarly, if we measure the first qubit and find it in the state $|1\rangle$, the second qubit will also be in the state $|1\rangle$, because:

$$\frac{1}{\sqrt{2}} (|00\rangle + |11\rangle) \xrightarrow{\text{measure 1st qubit to be } |1\rangle} |11\rangle. \quad (8.5)$$

This demonstrates perfect correlation between the measurement outcomes of the individual qubits in state $|\Phi^+\rangle$, measured along the z direction. Similarly, $|\Phi^-\rangle$ also exhibits perfect correlation, while $|\Psi^+\rangle$ and $|\Psi^-\rangle$ exhibit perfect anti-correlation. Later on (in § 9.8), you will discover that this strong correlation also exists in other directions.

8.2 *Bell Basis

In this section, we delve deeper into the mathematical properties of Bell states, specifically examining their orthonormality and completeness. The orthonormality of Bell states stems from the fact that the inner product of distinct Bell states is zero, whereas the inner product of a Bell state with itself is one. Their completeness, on the other hand, ensures that any two-qubit state can be represented as a linear combination of Bell states. As such, the set of Bell states forms an orthonormal basis for the two-qubit Hilbert space, a fundamental concept in quantum information theory and quantum computing. In the subsequent subsections, we'll explore these properties and their implications in more detail.

8.2.1 * Orthonormality and Completeness

The four Bell states are orthonormal because their inner products are either zero (when the states are different) or one (when the states are the same), which is proved in § 8.6. In the $|\beta_{ij}\rangle$ notation, this is expressed as:

$$\langle \beta_{ij} | \beta_{kl} \rangle = \delta_{ik} \delta_{jl}, \quad (8.6)$$

where δ_{ik} and δ_{jl} are Kronecker deltas and $i, j, k, l \in \{0, 1\}$.

Furthermore, $\{|\beta_{ij}\rangle\}$ is a complete basis set, as demonstrated by

$$\sum_{i,j \in \{0,1\}} |\beta_{ij}\rangle \langle \beta_{ij}| = I, \quad (8.7)$$

where I is the 4×4 identity matrix. This means that the Bell states span the entire two-qubit Hilbert space. Thus, the Bell states form an orthonormal basis for the two-qubit Hilbert space.



The set of Bell states serves as an orthonormal basis for two-qubit states, much like the set $|+\rangle, |-\rangle$ forms an orthonormal basis for single-qubit states.

8.2.2 * Two-Qubit States in Bell Basis

Any arbitrary two-qubit quantum state can be expressed as

$$|\psi\rangle = \alpha |00\rangle + \beta |01\rangle + \gamma |10\rangle + \delta |11\rangle. \quad (8.8)$$

Given that the Bell states form a complete basis for two-qubit states, we can also express $|\psi\rangle$ as a linear combination of these states:

$$|\psi\rangle = \sum_{i,j \in \{0,1\}} c_{ij} |\beta_{ij}\rangle, \quad (8.9)$$

where c_{ij} are complex coefficients. These coefficients must satisfy the normalization condition:

$$\sum_{i,j \in \{0,1\}} |c_{ij}|^2 = 1. \quad (8.10)$$

We can calculate the coefficients c_{ij} using the inner product between the Bell states and $|\psi\rangle$:

$$c_{ij} = \langle \beta_{ij} | \psi \rangle. \quad (8.11)$$

We can combine Eqs. 8.9 and 8.11 as a single equation using projectors $|\beta_{ij}\rangle \langle \beta_{ij}|$:

$$|\psi\rangle = \sum_{i,j \in \{0,1\}} |\beta_{ij}\rangle \langle \beta_{ij}| \psi \rangle. \quad (8.12)$$

In terms of the coefficients α, β, γ , and δ , the coefficients c_{ij} are:

$$c_{00} = \frac{1}{\sqrt{2}}(\alpha + \delta), \quad (8.13a)$$

$$c_{01} = \frac{1}{\sqrt{2}}(\beta + \gamma), \quad (8.13b)$$

$$c_{10} = \frac{1}{\sqrt{2}}(\beta - \gamma), \quad (8.13c)$$

$$c_{11} = \frac{1}{\sqrt{2}}(\alpha - \delta). \quad (8.13d)$$

This is equivalent to a basis-rotation using the “Bell matrix”:

$$\begin{bmatrix} c_{00} \\ c_{01} \\ c_{10} \\ c_{11} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix}. \quad (8.14)$$



The Bell state generator we’ll discuss in subsequent sections can be understood as a circuit that transforms the computational basis into the Bell basis. Conversely, the Bell state analyzer can be seen as performing the reverse transformation, from the Bell basis back to the computational basis.

Exercise 8.3 Derive the equations for the coefficients α , β , γ , and δ in terms of the amplitudes c_{ij} . This will provide the inverse relation of Eq. 8.13.

8.3 Bell State Creation

This section focuses on the practical creation of Bell states, which is a fundamental operation in quantum computing and quantum information theory. In an ideal quantum computer, Bell states can be readily generated using universal quantum gates like the Hadamard (H) and the controlled NOT (CNOT) gates. We will first investigate the basic methods of generating Bell states using these quantum gates. However, in practice, especially when working with systems like photons, we often resort to physical phenomena such as spontaneous parametric down-conversion (SPDC) to create entangled pairs of particles, which form the Bell states. Throughout this section, we will explore these methods and the underlying physics, providing a comprehensive understanding of Bell state creation in both theoretical and experimental scenarios.

8.3.1 Basic Bell State Generator

Bell states can be generated using qubit gates. For example, the Hadamard gate and the CNOT gate can be used together to create the Bell state Φ^+ from an initial state $|00\rangle$ as follows.

Apply a Hadamard gate (H) to the first qubit. This creates a superposition state:

$$(H \otimes I)|00\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle). \quad (8.15)$$

Apply a CNOT gate to the two-qubit state, with the first qubit as the control and the second qubit as the target. This entangles the qubits and generates the Bell state:

$$\text{CNOT } \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle) = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \quad (8.16)$$

All four Bell states can be created using qubit gates as in Fig. 8.1:

In	Out	
$ 00\rangle$	$ \beta_{00}\rangle \equiv \Phi^+\rangle = \frac{1}{\sqrt{2}}(00\rangle + 11\rangle)$	
$ 01\rangle$	$ \beta_{01}\rangle \equiv \Psi^+\rangle = \frac{1}{\sqrt{2}}(01\rangle + 10\rangle)$	
$ 10\rangle$	$ \beta_{10}\rangle \equiv \Phi^-\rangle = \frac{1}{\sqrt{2}}(00\rangle - 11\rangle)$	Bell-State Generator
$ 11\rangle$	$ \beta_{11}\rangle \equiv \Psi^-\rangle = \frac{1}{\sqrt{2}}(01\rangle - 10\rangle)$	

Figure 8.1: Bell State Creation using H and CNOT Gates



Clarification of Quantum Circuits and Math Formulas

In quantum circuit diagrams, time generally flows from left to right, and operations (gates) are applied to qubits as you move along this direction. However, when you represent the operations mathematically as matrices, the sequence of operations follows the convention of matrix multiplication, which is right-to-left.

This discrepancy can sometimes be confusing for those new to quantum computing. But it's an important aspect to understand for accurate interpretation of quantum circuits and corresponding mathematical expressions.

Exercise 8.4 Verify the generation of $|\Psi^+\rangle$, $|\Phi^-\rangle$, and $|\Psi^-\rangle$ in Fig. 8.1.

8.3.2 Generating All Bell States from $|00\rangle$

Adding X gates for the input state preparation, we can create all four Bell states directly from $|00\rangle$, as shown in Fig. 8.2. Remember that an X gate transforms a qubit from $|0\rangle$ to $|1\rangle$.

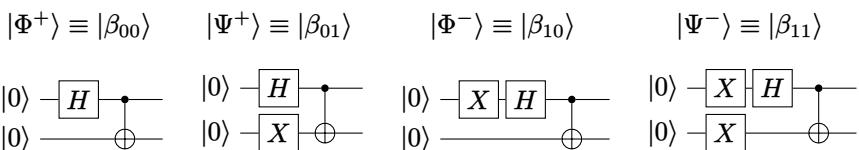


Figure 8.2: Bell State Creation from Initial State $|00\rangle$

Exercise 8.5 Verify the generation of $|\Phi^-\rangle$ and $|\Psi^-\rangle$ in Fig. 8.2 using operator algebra.

Alternatively, we can use the Z gate to generate a π phase shift (i.e., negative sign) in the output to create $|\Phi^-\rangle$ and $|\Psi^-\rangle$ as shown in Fig. 8.3. Remember that applying a Z gate to a qubit changes its $|1\rangle$ to $-|1\rangle$. (Applying the Z gate to either the first or the second qubit yields the same Bell states, as the difference is only in global phase which is not measurable in a quantum system.)

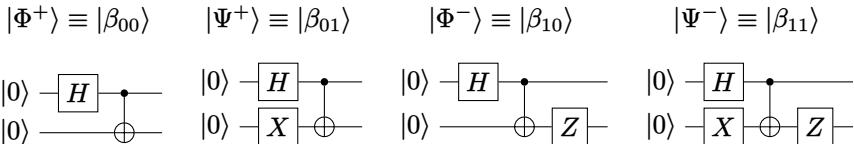


Figure 8.3: Bell State Creation Using Output Phase Shift

Exercise 8.6 Verify the generation of $|\Phi^-\rangle$ and $|\Psi^-\rangle$ in Fig. 8.3 using operator algebra.

8.3.3 Photon Pair Creation with SPDC

Direct application of Hadamard and CNOT gates to photons is not readily feasible. Instead, entangled photons are generated through spontaneous parametric down-conversion (SPDC) or other nonlinear optical processes. SPDC is a widely used method for producing pairs of entangled photons for a variety of quantum communication and computation experiments.

In SPDC, as illustrated in Fig. 8.4 a single high-energy photon passes through a nonlinear crystal and occasionally splits into two lower-energy photons that are entangled. The entangled photons can be created in various Bell states, depending on the experimental setup and the properties of the nonlinear crystal.

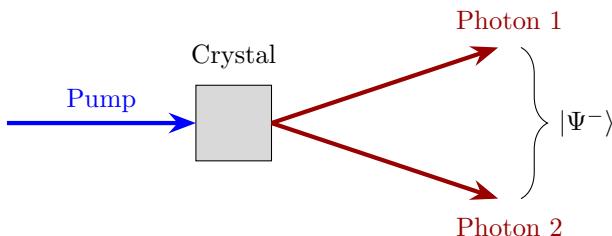


Figure 8.4: Photon Pair Creation with SPDC

For polarization-entangled photon pairs created by SPDC, the Bell state is

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}} (|\leftrightarrow\rangle_1 \otimes |\uparrow\rangle_2 - |\downarrow\rangle_1 \otimes |\leftrightarrow\rangle_2), \quad (8.17)$$

where $|\leftrightarrow\rangle_1$, $|\uparrow\rangle_1$, $|\leftrightarrow\rangle_2$ and $|\downarrow\rangle_2$ represent the horizontal and vertical polarizations of the two photons in the pair.

If we interpret $|\leftrightarrow\rangle$ as $|0\rangle$ and $|\uparrow\rangle$ as $|1\rangle$, then Eq. 8.17 yields the same $|\Psi^-\rangle$ as in the standard notation presented in Eq. 8.1.

8.3.4 * Boolean Representation

Using the Boolean representation introduced in § 7.3, the four Bell states can be succinctly expressed in a single equation:

$$|\beta_{xy}\rangle = \frac{1}{\sqrt{2}} (|0\rangle|y\rangle + (-1)^x |1\rangle|\bar{y}\rangle). \quad (8.18)$$

The Bell state creation circuit can be presented as in Fig. 8.5.

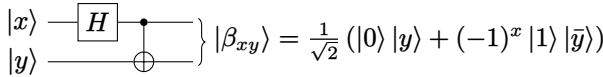


Figure 8.5: Boolean Representation of Bell States

8.4 Bell Measurement

This section delves into Bell measurements, which hold a pivotal position in quantum information processing due to their ability to distinguish between different Bell states. These measurements operate by projecting a two-qubit state onto the Bell basis, effectively resulting in one of the four Bell states. Such a process forms the core of numerous quantum protocols like quantum teleportation, superdense coding, and quantum key distribution. Moreover, we will take a look at how Bell measurements can be practically implemented using a quantum circuit known as a “Bell analyzer.”

8.4.1 Joint Measurement

The orthonormality of the Bell states (discussed in § 8.2) allows us to distinguish between different Bell states through measurements. If a quantum state is prepared in one of the Bell states, it can be unambiguously determined through a joint measurement on both qubits, known as a Bell measurement or a Bell state analysis.

A Bell measurement differs from the local measurements discussed in § 8.1.2 in that it is a type of joint quantum measurement (see § 6.2.2) performed on a two-qubit system. This implies the necessity of having both qubits at the same location or some form of quantum communication between the two qubits, which characterizes the non-local nature of Bell measurements.

If we perform a Bell measurement on a general two-qubit state $|\psi\rangle$ given by

$$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle \quad (8.19a)$$

$$= \sum_{i,j \in \{0,1\}} c_{ij} |\beta_{ij}\rangle, \quad (8.19b)$$

we will obtain one of the Bell states $|\beta_{ij}\rangle$ with probability $|c_{ij}|^2$, where $i, j \in \{0, 1\}$. The quantum state will then collapse into that Bell state. The coefficients c_{ij} are given by Eq. 8.13 in terms of α, β, γ , and δ .

Effectively, a Bell measurement can be thought of as projecting the two-qubit state $|\psi\rangle$ onto one of the four Bell states and performing the measurement in the Bell basis. The measurement operator for outcome $|\beta_{ij}\rangle$ is the projection operator:

$$M_{ij} = |\beta_{ij}\rangle\langle\beta_{ij}|. \quad (8.20)$$

The set of all measurement operators $\{M_{ij} | i, j \in \{0, 1\}\}$ forms a complete measurement set. That is, the sum of all measurement operators is the identity operator on the two-qubit Hilbert space (see Eq. 8.7). Compare these operators to the single-qubit local operators (e.g., $|0\rangle\langle 0| \otimes I$) used in § 8.1.2.

After a Bell measurement, all subsequent measurements on the unaltered system will yield the same outcome, as the state has been collapsed into the corresponding Bell state.

8.4.2 Basic Implementation: Bell Analyzer

A Bell measurement can be implemented by the circuit in Fig. 8.6. The Bell measurement circuit (or Bell analyzer) is the inverse of the Bell state generation circuit shown in Fig. 8.1. After the CNOT and Hadamard gates, we measure the two qubits in the computational basis. The first qubit gives us i and second j . Together we obtain one of four labels 00, 01, 10, or 11 indicating the measurement outcome $|\beta_{ij}\rangle$.

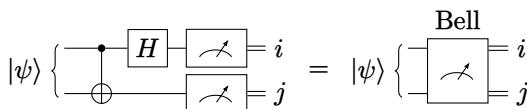


Figure 8.6: Bell Measurement

It's worth noting that the Bell analyzer can be utilized with any two-qubit state, not just Bell states. In the case of an arbitrary two-qubit state, the measurement results could be any combination of i and j , with each outcome occurring with a probability of $|c_{ij}|^2$. Here, i and j correspond to the indices of the Bell state to which the system collapses.

8.4.3 Post-Measurement State

In Fig. 8.6, we presume that our primary concern is the classical measurement results, i.e., the two-bit outcomes 00, 01, 10, or 11, and the resultant quantum state is disregarded. Nevertheless, there might be circumstances in which we want to use the Bell state $|\beta_{ij}\rangle$ that the system collapses into after the measurement in subsequent portions of the quantum circuit. In such cases, we would need to recreate that state, as illustrated in Fig. 8.7.

Note with this full implementation, the Bell measurement not only provides insight into the state of the system but also engenders a state change. If the qubits were initially not entangled, the measurement procedure will bring about their entanglement.

8.5 Bell State Conversion

Bell states can be converted from one to another using Pauli gates. In fact, this conversion can be achieved by applying a Pauli gate to one of the qubits locally. The ability to convert between Bell states forms an important aspect of quantum

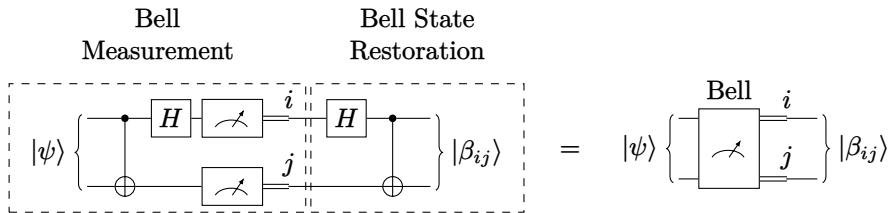


Figure 8.7: Bell Measurement with Post-Measurement State Restored

information processing and serves as a fundamental skill for anyone seeking to work in this field. This section sheds light on such conversions. It details converting $|\beta_{00}\rangle$ to other Bell states and vice versa.

8.5.1 Converting $|\beta_{00}\rangle$ to Other Bell States

Let's take $|\beta_{00}\rangle \equiv |\Phi^+\rangle$ as our reference state. Fig. 8.8 shows several ways to convert $|\beta_{00}\rangle$ to other Bell states.

$$\begin{array}{ccc}
 |\beta_{00}\rangle \left\{ \begin{array}{c} X \\ \hline \end{array} \right\} |\beta_{01}\rangle & |\beta_{00}\rangle \left\{ \begin{array}{c} Z \\ \hline \end{array} \right\} |\beta_{10}\rangle & |\beta_{00}\rangle \left\{ \begin{array}{c} Z \\ X \end{array} \right\} |\beta_{11}\rangle \\
 |\beta_{01}\rangle = (X \otimes I) |\beta_{00}\rangle & |\beta_{10}\rangle = (Z \otimes I) |\beta_{00}\rangle & |\beta_{11}\rangle = (Z \otimes X) |\beta_{00}\rangle \\
 \\
 |\beta_{00}\rangle \left\{ \begin{array}{c} \hline X \\ \hline \end{array} \right\} |\beta_{01}\rangle & |\beta_{00}\rangle \left\{ \begin{array}{c} \hline Z \\ \hline \end{array} \right\} |\beta_{10}\rangle & |\beta_{00}\rangle \left\{ \begin{array}{c} X \\ \hline Z \end{array} \right\} |\beta_{11}\rangle \\
 |\beta_{01}\rangle = (I \otimes X) |\beta_{00}\rangle & |\beta_{10}\rangle = (I \otimes Z) |\beta_{00}\rangle & |\beta_{11}\rangle = (ZX \otimes I) |\beta_{00}\rangle
 \end{array}$$

Figure 8.8: Converting $|\beta_{00}\rangle \equiv |\Phi^+\rangle$ to Other Bell States



The expression $|\beta_{11}\rangle = (Z \otimes X) |\beta_{00}\rangle$ is sometimes written in shorthand form as $|\beta_{11}\rangle = Z_1 X_2 |\beta_{00}\rangle$, where Z_1 refers to Z applied to the first qubit, and X_2 meaning X applied to the second qubit. This convention also applies to other similar expressions.

Exercise 8.7 Demonstrate $|\beta_{11}\rangle = (ZX \otimes I) |\beta_{00}\rangle$ which underlines the last conversion circuit in Fig. 8.8.

8.5.2 Converting Other Bell States to $|\beta_{00}\rangle$

Similarly, Fig. 8.9 shows several ways to convert other Bell states to $|\beta_{00}\rangle$. These are the inverse operations of the transformations in Fig. 8.8. You notice the forward conversion and reverse conversion are the same for the first five cases. This is because $X^2 = I$ and $Z^2 = I$. In the last case, the order of X and Z is different between forward and reverse conversion, because $XZZX = I$.

$$\begin{array}{lll}
|\beta_{01}\rangle \left\{ \begin{array}{c} \boxed{X} \\ \hline \end{array} \right\} |\beta_{00}\rangle & |\beta_{10}\rangle \left\{ \begin{array}{c} \boxed{Z} \\ \hline \end{array} \right\} |\beta_{00}\rangle & |\beta_{11}\rangle \left\{ \begin{array}{c} \boxed{Z} \\ \hline \boxed{X} \end{array} \right\} |\beta_{00}\rangle \\
|\beta_{00}\rangle = (X \otimes I) |\beta_{01}\rangle & |\beta_{00}\rangle = (Z \otimes I) |\beta_{10}\rangle & |\beta_{00}\rangle = (Z \otimes X) |\beta_{11}\rangle \\
\\
|\beta_{01}\rangle \left\{ \begin{array}{c} \hline \boxed{X} \end{array} \right\} |\beta_{00}\rangle & |\beta_{10}\rangle \left\{ \begin{array}{c} \hline \boxed{Z} \end{array} \right\} |\beta_{00}\rangle & |\beta_{11}\rangle \left\{ \begin{array}{c} \hline \boxed{Z} \quad \boxed{X} \\ \hline \end{array} \right\} |\beta_{00}\rangle \\
|\beta_{00}\rangle = (I \otimes X) |\beta_{01}\rangle & |\beta_{00}\rangle = (I \otimes Z) |\beta_{10}\rangle & |\beta_{00}\rangle = (XZ \otimes I) |\beta_{11}\rangle
\end{array}$$

Figure 8.9: Converting Other Bell States to $|\beta_{00}\rangle \equiv |\Phi^+\rangle$

Exercise 8.8 In Figs. 8.8 and 8.9, we have adopted $|\beta_{00}\rangle \equiv |\Phi^+\rangle$ as the standard Bell state, following the usual convention in quantum computing. However, in quantum communication, photon pairs generated via Spontaneous Parametric Down-Conversion (SPDC) often exhibit the state $|\beta_{11}\rangle \equiv |\Psi^-\rangle$. Recreate comparable figures using $|\Psi^-\rangle$ as the reference state.

8.5.3 Additional Conversions Related to $|\beta_{11}\rangle$

Some acute readers may have discovered that some combinations for $|\beta_{11}\rangle$ are missing in Figs. 8.8 and 8.9. Well, some combinations introduce a global phase, which in practical sense, is not a problem. These are shown in Fig. 8.10.

$$\begin{array}{ll}
|\beta_{11}\rangle \left\{ \begin{array}{c} \boxed{X} \\ \hline \boxed{Z} \end{array} \right\} - |\beta_{00}\rangle & |\beta_{11}\rangle \left\{ \begin{array}{c} \boxed{X} \quad \boxed{Z} \\ \hline \end{array} \right\} - |\beta_{00}\rangle \\
|\beta_{00}\rangle = -(X \otimes Z) |\beta_{11}\rangle & |\beta_{00}\rangle = -(ZX \otimes I) |\beta_{11}\rangle \\
\\
|\beta_{00}\rangle \left\{ \begin{array}{c} \hline \boxed{X} \\ \hline \boxed{Z} \end{array} \right\} - |\beta_{11}\rangle & |\beta_{00}\rangle \left\{ \begin{array}{c} \hline \boxed{Z} \quad \boxed{X} \\ \hline \end{array} \right\} - |\beta_{11}\rangle \\
|\beta_{11}\rangle = -(X \otimes Z) |\beta_{00}\rangle & |\beta_{11}\rangle = -(XZ \otimes I) |\beta_{00}\rangle
\end{array}$$

Figure 8.10: Additional Conversions Related to $|\beta_{11}\rangle \equiv |\Psi^-\rangle$

8.6 *Deferred Proofs

Equation 8.11 can be proved as follows:

$$\langle \beta_{ij} | \psi \rangle = \langle \beta_{ij} | \sum_{k,l \in \{0,1\}} c_{kl} |\beta_{kl} \rangle \rangle \quad (8.21a)$$

$$= \sum_{k,l \in \{0,1\}} c_{kl} \langle \beta_{ij} | \beta_{kl} \rangle \quad (8.21b)$$

$$= \sum_{k,l \in \{0,1\}} c_{kl} \delta_{ik} \delta_{jl} \quad (8.21c)$$

$$= c_{ij}. \quad (8.21d)$$

Now, let's prove Eq. 8.6, the orthonormality of the Bell states.

For normalization, let's calculate the inner product of each Bell state with itself:

$$\langle \beta_{00} | \beta_{00} \rangle = \frac{1}{2} (\langle 00|00 \rangle + \langle 00|11 \rangle + \langle 11|00 \rangle + \langle 11|11 \rangle) = \frac{1}{2} (1 + 0 + 0 + 1) = 1,$$

$$\langle \beta_{01} | \beta_{01} \rangle = \frac{1}{2} (\langle 00|00 \rangle - \langle 00|11 \rangle - \langle 11|00 \rangle + \langle 11|11 \rangle) = \frac{1}{2} (1 - 0 - 0 + 1) = 1,$$

$$\langle \beta_{10} | \beta_{10} \rangle = \frac{1}{2} (\langle 01|01 \rangle + \langle 01|10 \rangle + \langle 10|01 \rangle + \langle 10|10 \rangle) = \frac{1}{2} (1 + 0 + 0 + 1) = 1,$$

$$\langle \beta_{11} | \beta_{11} \rangle = \frac{1}{2} (\langle 01|01 \rangle - \langle 01|10 \rangle - \langle 10|01 \rangle + \langle 10|10 \rangle) = \frac{1}{2} (1 - 0 - 0 + 1) = 1.$$

For orthogonality, let's calculate the inner product of distinct Bell states:

$$\langle \beta_{01} | \beta_{00} \rangle = \frac{1}{2} (\langle 01|00 \rangle + \langle 01|11 \rangle + \langle 10|00 \rangle + \langle 10|11 \rangle) = \frac{1}{2} (0 + 0 + 0 + 0) = 0,$$

$$\langle \beta_{01} | \beta_{10} \rangle = \frac{1}{2} (\langle 01|00 \rangle - \langle 01|11 \rangle + \langle 10|00 \rangle - \langle 10|11 \rangle) = \frac{1}{2} (0 - 0 + 0 - 0) = 0,$$

$$\langle \beta_{01} | \beta_{11} \rangle = \frac{1}{2} (\langle 01|01 \rangle - \langle 01|10 \rangle + \langle 10|01 \rangle - \langle 10|10 \rangle) = \frac{1}{2} (1 - 0 + 0 - 1) = 0,$$

$$\langle \beta_{10} | \beta_{00} \rangle = \frac{1}{2} (\langle 00|00 \rangle + \langle 00|11 \rangle - \langle 11|00 \rangle - \langle 11|11 \rangle) = \frac{1}{2} (1 + 0 - 0 - 1) = 0,$$

$$\langle \beta_{10} | \beta_{11} \rangle = \frac{1}{2} (\langle 00|01 \rangle - \langle 00|10 \rangle + \langle 11|01 \rangle - \langle 11|10 \rangle) = \frac{1}{2} (0 - 0 + 0 - 0) = 0,$$

$$\langle \beta_{00} | \beta_{11} \rangle = \frac{1}{2} (\langle 00|01 \rangle - \langle 00|10 \rangle + \langle 11|01 \rangle - \langle 11|10 \rangle) = \frac{1}{2} (0 - 0 + 0 - 0) = 0.$$

All the inner products of distinct Bell states are equal to 0, which indicates that the Bell states are orthogonal to each other.

Exercise 8.9 Prove the completeness of Bell basis, i.e., Eq. 8.7.

8.7 *Generalization: GHZ States

Bell states, or EPR pairs, involve two entangled qubits. Generalizing Bell states to more than two qubits can lead to more complex types of entanglement. There are numerous ways to entangle more than two qubits, and these lead to different classes of multi-qubit entangled states.

GHZ (Greenberger-Horne-Zeilinger) states are a class of multi-qubit states that demonstrate stronger forms of quantum entanglement than can be seen with two qubits. The simplest GHZ state involves three qubits and is given by the superposition

$\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$, where each term in the superposition involves all qubits being in the same state.

When we increase the number of qubits, we encounter more types of multipartite entangled states, such as the W states, e.g., $\frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$. It's worth noting that these different classes of entangled states cannot always be converted into each other without some probability of failure, which suggests they demonstrate fundamentally different types of quantum correlation.

* Further Exploration

The concept of GHZ states, as well as their applications in quantum information theory and quantum computation, extends beyond the basic explanation given in this text. For those seeking to deepen their understanding, the following topics are suggested for further exploration:

1. Going Beyond Bell's Theorem [48]: In this seminal paper, Greenberger, Horne, and Zeilinger introduce the concept of GHZ states, which exemplify a more profound form of quantum entanglement than Bell states.
2. Quantum Computation and Quantum Information [7]: This comprehensive textbook by Nielsen and Chuang covers the entire field of quantum information, including the detailed theory of multi-qubit entangled states such as GHZ and W states.
3. Observation of Three-Photon Greenberger-Horne-Zeilinger Entanglement [27]: In this experimental paper, Bouwmeester et al. demonstrate the creation of a three-qubit GHZ state, putting the theoretical predictions into practice.

8.8 Summary and Conclusions

Bell States and Quantum Entanglement

Throughout this chapter, we have ventured into the realm of Bell states and their profound implications. By studying Bell states, we were able to demystify complex quantum phenomena like quantum entanglement and to understand them in a precise mathematical framework. We emphasized the indispensable nature of Bell states in the realm of quantum computation and information processing.

Theoretical Foundations and Practical Implementations

We subsequently dissected the properties of Bell states, establishing their role as the quintessential representation of two-qubit entanglement. The concepts of orthonormality and completeness of Bell states were meticulously explored, highlighting their significance in quantum computing and information.

Our expedition into the practical creation of Bell states encompassed both theoretical constructs, employing quantum gates, and real-world scenarios utilizing physical phenomena like spontaneous parametric down-conversion (SPDC).

Measurement and Conversion

The chapter further unveiled the intricacies of Bell measurements, which are essential for many quantum protocols. Their ability to discern between different Bell states and the projection mechanism onto the Bell basis was elaborated upon. Moreover,

the concept of Bell state conversion was introduced, showcasing the flexibility and adaptability of these entangled states in quantum operations.

Generalizations and Future Directions

In closing the chapter, we introduced the reader to the more general entangled states such as the GHZ states. These states serve as a natural extension of the Bell states, embracing more than two qubits. Such generalizations enrich our understanding of multi-qubit entanglement and open doors for further research and applications.

Upcoming Topics

In the succeeding chapter, we will venture deeper into the enigmatic domain of quantum entanglement, specifically focusing on Bell inequalities. These inequalities will be our tool to probe the disparity between quantum mechanics and the principles of local realism. By delving into experimental tests that validate quantum entanglement and the consequent negation of local hidden variables, we shall further unravel the mysteries of the quantum realm. Additionally, we will explore the pragmatic applications of quantum entanglement, particularly in the domains of secure communication and cryptographic protocols.

Problem Set 8

8.1 A general form of two-qubit Bell states is

$$|\Phi\rangle = \frac{1}{\sqrt{2}}(|ab\rangle + |cd\rangle),$$

with $\langle a|c\rangle = \langle b|d\rangle = 0$.

(a) Prove that $|\Phi\rangle$ is not a product state.

(b) Confirm this general form encompasses the definitions in Eq. 8.1.

8.2 (a) Using the general form of Bell states in Problem 8.1, show that if $|\Psi\rangle$ and $|\Phi\rangle$ are two Bell states shared by Alice and Bob, then there is a unitary operator local to Alice (i.e., of the form $U \otimes I$) which takes $|\Psi\rangle$ to $|\Phi\rangle$:

$$(U \otimes I) |\Psi\rangle = |\Phi\rangle.$$

Similarly, there is such a unitary operator local to Bob. Thus, all Bell states are equivalent in this sense.

(b) Find U for the specific conversions in § 8.5.

8.3 Derive the unitary matrix U_β for the basis change from the computational basis to the Bell basis. This transformation is important in quantum computing for encoding and decoding information in Bell states. Specifically, find U_β such that

$$[|\beta_{00}\rangle, |\beta_{01}\rangle, |\beta_{10}\rangle, |\beta_{11}\rangle] = U_\beta [|\Psi\rangle, |\Phi\rangle, |\Phi\rangle, |\Psi\rangle],$$

where each ket ($| \dots \rangle$) represents a column vector.

Relate U_β to the specific cases in § 8.3.

8.4 For some qubit platforms, it is more natural to implement the Bell state

$$|\Phi_i^+\rangle = \frac{1}{\sqrt{2}} (|00\rangle + i|11\rangle)$$

than the standard Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)$.

Find the other three Bell states ($|\Phi_i^-\rangle$, $|\Psi_i^+\rangle$, $|\Psi_i^-\rangle$) such that, together with $|\Phi_i^+\rangle$, the four states form a complete orthonormal basis.

8.5 Consider the three-qubit state given by:

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}} (\alpha|0\rangle + \beta|1\rangle) \otimes (|00\rangle + |11\rangle).$$

Assume we measure the first two qubits in the Bell basis $|\beta_{ij}\rangle$ (where $i, j \in \{0, 1\}$).

- (a) What is the probability for each of the Bell states $|\beta_{ij}\rangle$ to occur?
- (b) If the measurement yields $|\beta_{ij}\rangle$, what is the state of the third qubit after the measurement?

8.6 Consider the four-qubit state given by:

$$|\Psi_0\rangle = \frac{1}{2} (|0000\rangle + |0101\rangle + |1010\rangle + |1111\rangle).$$

Assume we measure the first two qubits in the Bell basis $|\beta_{ij}\rangle$ (where $i, j \in \{0, 1\}$).

- (a) What is the probability for each of the Bell states $|\beta_{ij}\rangle$ to occur?
- (b) If the measurement yields $|\beta_{ij}\rangle$, what is the state of the last two qubits after the measurement?

8.7 * Consider the Bell state

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}} (|01\rangle - |10\rangle).$$

Let \mathbf{n} and \mathbf{m} be unit vectors in \mathbb{R}^3 . Calculate the expectation value

$$E(\mathbf{n}, \mathbf{m}) = \langle \Psi^- | (\mathbf{n} \cdot \boldsymbol{\sigma}) \otimes (\mathbf{m} \cdot \boldsymbol{\sigma}) | \Psi^- \rangle$$

where $\boldsymbol{\sigma}$ represents the vector of Pauli matrices. Show that

$$E(\mathbf{n}, \mathbf{m}) = -n_1 m_1 - n_2 m_2 - n_3 m_3 = -\mathbf{n} \cdot \mathbf{m} = -\cos \theta_{\mathbf{n}, \mathbf{m}}.$$

8.8 * Multipartite Bell States

Consider the extension of the Bell state concept to a system of four qubits, or equivalently, a pair of qudits with $d = 4$. We define the uniform Bell state for such a system as follows:

$$|\beta_{0000}\rangle = \frac{1}{2} \sum_{i,j \in \{0,1\}} |ij\rangle \otimes |ij\rangle = \frac{1}{2}(|0000\rangle + |0101\rangle + |1010\rangle + |1111\rangle).$$

This state can also be represented in decimal notation as:

$$|\beta_0\rangle = \frac{1}{2} \sum_{i=0}^3 |i\rangle \otimes |i\rangle = \frac{1}{2}(|0\rangle + |5\rangle + |10\rangle + |15\rangle),$$

with $|i\rangle$ now ranging over the set $\{0, 1, 2, 3\}$, corresponding to the decimal equivalent of the binary states.

The other Bell states can be generated using the transformation:

$$|\beta_{ijkl}\rangle = U_{ijkl} |\beta_{0000}\rangle \quad (i, j, k, l \in \{0, 1\}),$$

with U_{ijkl} defined as the application of specific Pauli and identity operators on the base Bell state. The operators are:

$$\begin{array}{llll} U_{0000} = II\,II & U_{0001} = IX\,II & U_{0010} = IZ\,II & U_{0011} = IY\,II \\ U_{0100} = XI\,II & U_{0001} = XX\,II & U_{0010} = XZ\,II & U_{0011} = XY\,II \\ U_{0100} = ZI\,II & U_{0101} = ZX\,II & U_{0110} = ZZ\,II & U_{0111} = ZY\,II \\ U_{1100} = YI\,II & U_{1101} = YX\,II & U_{1110} = YZ\,II & U_{1111} = YY\,II \end{array}$$

Your tasks are as follows:

- (a) Show that these Bell states are orthonormal, i.e.,

$$\langle \beta_{ijkl} | \beta_{i'j'k'l'} \rangle = \delta_{ii'} \delta_{jj'} \delta_{kk'} \delta_{ll'}.$$

- (b) Demonstrate that these Bell states form a complete basis for the space of four qubits, i.e.,

$$\sum_{i,j,k,l \in \{0,1\}} |\beta_{ijkl}\rangle \langle \beta_{ijkl}| = I.$$

- (c) Extend this construction to describe Bell states for systems comprising $2n$ qubits. (In the above example $n = 2$.)

9. Entanglement and Bell Inequalities

Contents

9.1	Classical Correlation vs. Quantum Entanglement	218
9.1.1	Classical Correlations	218
9.1.2	Quantum Entanglement	219
9.2	The EPR Paradox	221
9.2.1	Hidden Variable Theory, Realism, and Locality	222
9.3	Bell Inequalities	223
9.3.1	Bell-CHSH Inequality	223
9.4	Bell-CHSH Inequality with Classical Correlation	225
9.4.1	Balls with Contrasting Colors	225
9.4.2	Gray Balls with Shared Ink	226
9.4.3	Adaptable Intelligent Balls	226
9.5	Bell-CHSH Inequality with Quantum Entanglement	227
9.5.1	Violation of the Bell CHSH Inequality: A Special Case	227
9.5.2	Violation of the Bell CHSH Inequality: General Case	229
9.6	Experimental Verification	230
9.6.1	* Basic Experimental Setup	230
9.6.2	Early Experiments	231
9.6.3	Loophole-free Measurements	232
9.6.4	Key Experimental Findings and Implications	232
9.6.5	2022 Nobel Prize in Physics	233
9.7	The No-Communication Theorem	234
9.7.1	Myths and Misconceptions	234
9.8	* Derivation of Bell-CHSH Quantity for Bell States	235
9.8.1	* Derivation via Measurement Bases	235
9.8.2	* Derivation Using Pauli Operators	238
9.9	* Further Exploration	240
9.10	Summary and Conclusions	241
	Problem Set 9	242

In this chapter, we are going to embark on an exploration of the captivating world of quantum entanglement and Bell inequalities. Quantum entanglement is a phenomenon that challenges our conventional wisdom, especially the notion of local realism, which postulates that physical systems have predetermined properties and cannot interact faster than light.

One of the most famous examples that highlight the puzzling nature of quantum entanglement is the century-old Einstein-Podolsky-Rosen (EPR) paradox. This thought experiment raised questions about how complete and accurate quantum mechanics is, and sparked many discussions among scientists.

In recent years, scientists have tested quantum entanglement using Bell inequalities. These inequalities are tools used to distinguish between predictions made by quantum mechanics and the principles of local realism. Experiments that have confirmed quantum entanglement have greatly expanded our knowledge of the quantum world.

Beyond being a mere intellectual puzzle, quantum entanglement boasts tangible applications, becoming integral to emerging technologies. The E91 protocol, for instance, harnesses entanglement for secure communication through cryptographic keys and employs Bell inequalities to detect eavesdropping.

This chapter will thoroughly examine the historical background, theories, and experimental achievements of quantum entanglement and Bell inequalities. In the next chapter, we will study their real-world applications.

9.1 Classical Correlation vs. Quantum Entanglement

Quantum entanglement is a fundamental phenomenon in quantum mechanics where particles become correlated in such a way that the state of one particle cannot be described independently of the state of the other, even when they are separated by large distances. The physics underlying quantum entanglement is distinct from classical correlation.

To illustrate this distinction, consider a machine that emits pairs of balls in opposite directions, as depicted in Fig. 9.1. Envisage two individuals, Alice and Bob, each catching one of these balls. The balls are arranged such that when Bob catches a black ball, he knows that Alice must have caught a white one, and vice versa. This correlation can manifest in both classical and quantum realms but arises for different reasons.

9.1.1 Classical Correlations

In the classical domain, correlations can occur in various scenarios:

- **Balls with contrasting colors:** The machine dispenses pairs of balls with contrasting colors that are predetermined. In this scenario, Alice and Bob always catch balls of opposite colors because they were already arranged that way. This is similar to the predetermined coin-pair scenario: If a pair of coins is tossed such that when one shows heads, the other must show tails.

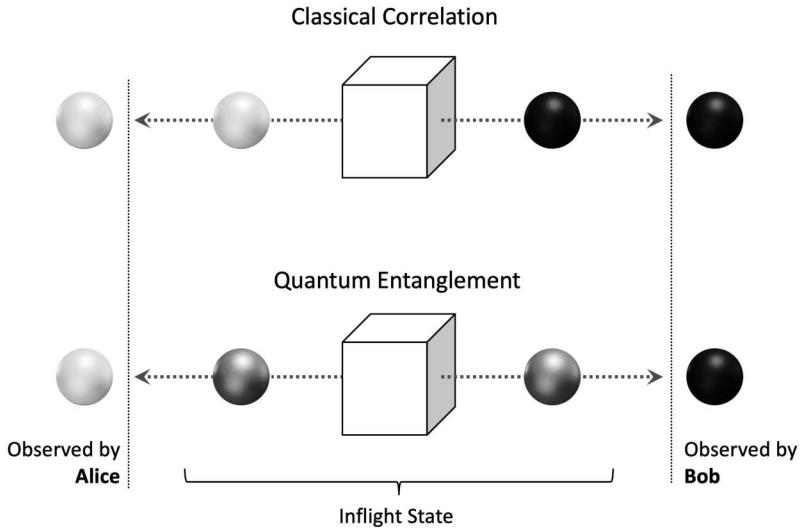


Figure 9.1: Classical Correlation versus Quantum Entanglement

- Gray balls with shared ink: Initially, the balls contain an equal mixture of black and white pigments, hence appearing gray. Upon observation, the pigments separate, flowing to one of the balls, with the direction of the flow being random. As a result, one ball randomly turns black, and the other turns white. In this scenario, the colors are not predetermined, and a transfer of pigment between the balls is necessary for them to assume their respective colors.
- Communicating intelligent balls: Consider two balls capable of sensing their surroundings and communicating wirelessly. They can coordinate to display different colors based on the environment (e.g., who catches them and how they are caught).



The examples above illustrate anti-correlation. Correlation can also be direct. For instance, with fair coins, if Alice's coin shows heads, so does Bob's, and the same applies for tails.

9.1.2 Quantum Entanglement

Transitioning to the quantum realm, let's consider the two balls as particles, or qubits. The scenario in Fig. 9.1 can occur if the qubits are in a Bell state, represented as $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. Here, we associate the white color with state $|0\rangle$ and black with state $|1\rangle$.

When Alice measures her qubit, she employs the projection operator

$$M_{0a} = |0\rangle\langle 0| \otimes I, \quad (9.1)$$

which collapses the Bell state to $M_{0a}|\Psi^-\rangle = |01\rangle$. (The normalization factor is not relevant here and is therefore omitted.) Subsequently, Bob's qubit is in state $|1\rangle$.

Likewise, if Bob measures his qubit first and observes $|0\rangle$, he uses the projection operator

$$M_{0b} = I \otimes |0\rangle\langle 0|, \quad (9.2)$$

which collapses the Bell state to $M_{0b} |\Psi^-\rangle = |10\rangle$. At this point, Alice's qubit is in state $|1\rangle$.

This demonstrates that the measurement outcomes of Alice and Bob are anti-correlated. However, a key difference from classical correlation lies in the quantum measurement principle. The states of the individual qubits in the entangled pair are not defined until they are measured. This is shown in Fig. 9.1, where the balls are gray before being caught, symbolizing their undefined states. When Alice measures her qubit, she obtains either $|0\rangle$ or $|1\rangle$ with equal probability, but whichever she gets, Bob will always observe the opposite. This resembles the “grey balls with shared ink” classical scenario, except that no communication or transfer of matter is needed between the qubits.

Similar logic applies to other Bell states. For $|\Phi^+\rangle$ and $|\Phi^-\rangle$, the measurement outcomes are correlated; for $|\Psi^+\rangle$ and $|\Psi^-\rangle$, they are anti-correlated.

The above discussion on entanglement is quite abstract. You might be wondering, how are these “gray balls” created in the physical world? In quantum computing, entanglement between qubits is achieved through a variety of physical implementations, each exhibiting the behavior described earlier. The following examples should help provide a more tangible understanding.

- Photon Pairs Generated With SPDC: Spontaneous parametric down-conversion (SPDC) is a process where a single photon decays into two entangled photons with lower energy. These entangled photon pairs show quantum correlations and can be sent to distant locations.
- Entangled Atomic States: Quantum entanglement can also include other particles, such as atoms. Entangled atomic states can be produced by changing the internal energy levels of atoms using techniques like stimulated Raman adiabatic passage (STIRAP).
- Superconducting Qubits: In quantum computing, superconducting qubits can be entangled through interactions controlled by microwave resonators or other coupling methods. This enables the creation of quantum logic gates and algorithms.
- Trapped Ions: Trapped ion systems create entangled states by controlling the motion and internal energy levels of ions held in electromagnetic traps. These systems are used for quantum computing and simulation.
- NV Centers: Nitrogen-vacancy (NV) centers in diamond are defect structures that can hold electron spins for a long time. These spins can be entangled using methods such as resonant microwave driving and optical control.
- Quantum Dots: Quantum dots are tiny semiconductor structures that can hold and control individual electrons or excitons. Entangled states can be created in quantum dots through methods like parametric fluorescence or electron spin interactions.

9.2 The EPR Paradox

Quantum mechanics posits that particles were not predetermined to be in a certain state but were actually in a superposition of multiple states until observed. For instance, in the context of the example given earlier, before observation, each particle could be thought of as being both black and white at the same time (analogous to being gray). The moment they are observed, they randomly collapse into one of the states, instantaneously determining the state of the other particle, no matter how far apart they are.

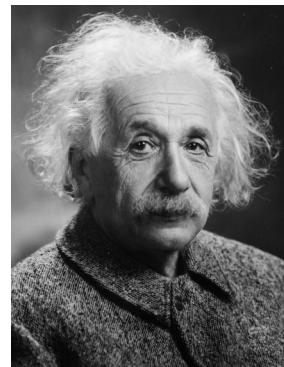
This seems to contradict our classical intuitions about cause and effect and the nature of reality. How can one particle be influenced by an observation made on another particle at a distance without any signal passing between them? According to relativity, a signal cannot travel faster than light, but quantum mechanics doesn't seem to require a signal for this entangled behavior. Moreover, how can it be that the outcomes of our observations are random and probabilistic?

Albert Einstein, along with his colleagues Boris Podolsky and Nathan Rosen, found this counterintuitive. In 1935, they argued that quantum mechanics does not provide a complete description of reality. This argument came to be known as the EPR paradox, named after the initials of the three scientists.

“ I cannot seriously believe in quantum theory because the theory cannot be reconciled with the idea that physics should represent a reality in time and space, free from *spooky actions at a distance*.

Do you really believe that *the moon isn’t there when nobody looks?*

Quantum mechanics is very worthy of regard. But an inner voice tells me that this is not yet the right track. The theory yields much, but it hardly brings us closer to the Old One’s secrets. I, in any case, am convinced that *He does not play dice.*”



Albert Einstein

Credit: Library of Congress

Interpretations and Philosophical Implications of Quantum Mechanics

Quantum mechanics is both an experimentally verified and mathematically rigorous theory, yet it opens avenues for philosophical debates. It questions our understanding of reality, the role of observers, and concepts like determinism and causality. Prominent physicists like Niels Bohr and Richard Feynman have noted the inherent challenges in fully comprehending the theory. There are currently several interpretations of quantum mechanics leading to different philosophical implications.

Copenhagen Interpretation: The Copenhagen interpretation, which we have implicitly followed so far, postulates that after a quantum measurement, the state

vector collapses to the measured eigenstate. In this view, quantum mechanics is fundamentally a theory of probability amplitudes, allowing for phenomena like superposition and entanglement.

Realist Interpretation: Contrastingly, the realist interpretation suggests that quantum mechanics is deterministic, and any uncertainties arise due to hidden variables. Albert Einstein was a strong advocate of this view, encapsulated in his phrase “God does not play dice.” However, modern experiments, including those recognized by the 2022 Nobel Prize, have largely ruled out this interpretation.

Other Interpretations: Various other interpretations exist, including many-worlds, information-based interpretations, and decoherence-based views. Each offers a unique perspective on the underlying principles of quantum mechanics.

9.2.1 Hidden Variable Theory, Realism, and Locality

EPR proposed an alternative to standard quantum mechanics that employs hidden variables. They assumed that particles always contained information, which might be hidden, about their states and that this information determined their states even before being observed. Essentially, this information caused the correlation between particles.

In this context, “hidden variables” refer to unobservable factors hypothesized to exist to account for the behavior of quantum systems in a deterministic way. The concept of hidden variables is associated with local realism, which encompasses two philosophical ideas: realism and locality.

Realism: This is the belief that objects possess definite properties and values even when they are not being observed or measured. In the context of hidden variables, realism implies that the properties of a quantum system (such as position, spin, or polarization) have well-defined values at all times, even when not being measured. These values are thought to be determined by hidden variables.

Locality: This refers to the idea that physical processes occurring at one location do not depend on the properties of objects at other locations that are spacelike separated (i.e., separated by a distance greater than the speed of light multiplied by the time it takes for information to travel between them). In the context of hidden variables, locality means that the outcomes of measurements on spatially separated particles are independent of each other and are determined solely by the local hidden variables associated with each particle.



In § 9.1.1, two examples - balls with contrasting colors and predetermined coin pairs - adhere to the principle of local realism. However, the other two examples - gray balls with shared ink and communicating intelligent balls - violate this principle.

However, the notion of local hidden variables faced challenges from Bell’s inequalities. Proposed by physicist John Bell in 1964, Bell’s inequalities are a set of mathematical inequalities that, if violated by experimental data, indicate that the underlying theory cannot be explained by local hidden variables. The violation of Bell’s inequalities by experimental data suggests that entangled quantum systems exhibit correlations that cannot be explained by classical physics.

Subsequent experiments consistently demonstrated violations of Bell's inequalities, suggesting that the behavior of entangled particles does not align with the concept of local hidden variables. This, in turn, provides strong support for the predictions of quantum mechanics and has profound implications for our understanding of the fundamental nature of reality.

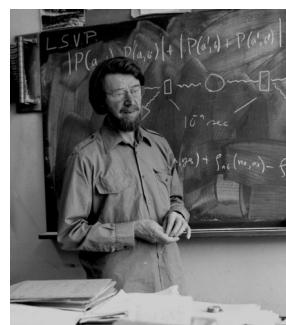
Key Takeaways

Classical correlations and quantum entanglement both involve connections between separate systems, but the nature of these connections is fundamentally different. Classical correlations under the assumption of local realism can be explained by predetermined states, whereas quantum entanglement involves an inherent, non-classical correlation that exists even without prior conditions or communication.

In the next section, we will explore in more detail how Bell's inequalities have shaped the modern understanding of quantum mechanics and its intriguing properties.

9.3 Bell Inequalities

Bell inequalities provide a mathematical framework that allows experiments to differentiate between the predictions of local hidden variable theories and those of quantum mechanics in the context of the EPR paradox. There are several forms of Bell inequalities, each applicable to various scenarios and types of observables. The original Bell inequality was defined for dichotomic observables (i.e., observables with two possible outcomes). The Bell-CHSH inequality is a specific case derived by John Clauser, Michael Horne, Abner Shimony, and Richard Holt in 1969, which is an extension of the original Bell inequality and is designed for scenarios with two entangled particles and dichotomic observables.



John Bell

Credit: CERN Science Photo

In this text, we will focus on the Bell-CHSH inequality, as it is one of the most widely studied and experimentally tested forms of Bell inequalities. It serves as a cornerstone in our understanding of the fundamental differences between classical and quantum correlations, and it provides a clear demonstration of the nonlocal and counter-intuitive nature of quantum entanglement.

9.3.1 Bell-CHSH Inequality



First time readers and undergraduate students may skip the following derivation until Eq. 9.7.

1 Simplified Form

The Bell-CHSH inequality is derived based on the assumptions of local realism, using classical probability theory. The derivation starts by defining measurement settings

for a pair of objects (colored balls, or particles in the hidden variable theory), A sent to Alice and B sent to Bob.

Each object can be measured along two different measurement settings or bases, such as polarization directions, represented by a and a' for Alice, and b and b' for Bob. Assume that Alice randomly chooses between a and a' , and Bob randomly chooses between b and b' , to measure each pair of objects.

Let's define $A(a)$ and $B(b)$ as the outcomes of the measurements for Alice and Bob, respectively, with respect to the settings specified by a and b . Similarly, $A(a')$ and $B(b')$ for a' and b' .

We assume the outcomes of these measurements are either $+1$ or -1 , that is:

$$A(a), A(a'), B(b), B(b') \in \{-1, 1\}. \quad (9.3)$$

Now, let's consider the following quantities:

$$\begin{aligned} & A(a) [B(b) - B(b')] \\ & A(a') [B(b) + B(b')] \end{aligned}$$

Since the outcomes of the measurements are either $+1$ or -1 , depending on if $B(b)$ and $B(b')$ have the same sign or opposite signs, one of the above quantities is 0 while the other is bounded to $[-2, 2]$.

Thus, the sum of the above quantities is also bounded to $[-2, 2]$:

$$-2 \leq A(a) [B(b) - B(b')] + A(a') [B(b) + B(b')] \leq 2. \quad (9.4)$$

2 Statistically Verifiable Version

However, Inequality 9.4 is not directly verifiable. This is because for each pair of objects, Alice randomly chooses between a and a' to measure, and Bob between b and b' . There are four combinations, and they need to measure at least four pairs to test the inequality thoroughly. To address this, we can derive a statistically verifiable version as follows.

Let's introduce λ , the “hidden variables,” which encompass all factors, such as initial conditions and environmental influences, that determine the states and properties of the objects. These factors may be inherently random. In classical probability theory, this is represented as a probability distribution function $P(\lambda)$. It is also assumed that the choice of measurement setting (such as a or a' for Alice) is independent of the hidden variables and does not influence the properties of the objects being measured, thus there is a single $P(\lambda)$ function.

Now $A(a)$, $A(a')$, $B(b)$, and $B(b')$ all depend on λ , and we denote them as $A(a, \lambda)$, $A(a', \lambda)$, $B(b, \lambda)$, and $B(b', \lambda)$.

The expected value of the product of two measurement outcomes, which describes the correlation of measurements under two settings, is given by:

$$E(a, b) = \sum_{\lambda} P(\lambda) A(a, \lambda) B(b, \lambda). \quad (9.5)$$

These expected values can be obtained through repeated measurements on many identically prepared pairs of particles and averaging the outcomes.

Now, we sum inequality (9.4) over the hidden variable λ with its probability distribution $P(\lambda)$:

$$-2 \leq \sum_{\lambda} P(\lambda) [A(a, \lambda) [B(b, \lambda) - B(b', \lambda)] + A(a', \lambda) [B(b, \lambda) + B(b', \lambda)]] \leq 2. \quad (9.6)$$

Using the definition of the expected value, we can rewrite the inequality as:

$$-2 \leq S \leq 2, \text{ or, } |S| \leq 2, \quad (9.7)$$

which is the Bell-CHSH inequality, where

$$S = E(a, b) - E(a, b') + E(a', b) + E(a', b') \quad (9.8)$$

is referred to as the Bell-CHSH quantity.

Points to Note

Quantum mechanics is not involved in the derivation of the Bell-CHSH inequality. It is purely based on classical probability theory and the hidden variable (i.e., local realism) assumption. As such, the inequality sets constraints on the correlations between measurements of two particles under any local hidden variable theory. If the inequality is violated, it suggests that the system in question does not adhere to local realism but instead exhibits the nonlocal correlations that are characteristic of quantum entanglement.

9.4 Bell-CHSH Inequality with Classical Correlation

9.4.1 Balls with Contrasting Colors

To build intuition, let's start with a classical analogy using "balls with contrasting colors." We will consider pairs of balls where each ball is either black or white. We define $+1$ for white, and -1 for black.

Imagine that Alice and Bob each receive one ball from a pair, and they have some method to measure or analyze the colors of the balls. We denote Alice's measurement settings as a and a' , and Bob's settings as b and b' .

Assume that Alice and Bob randomly choose between their respective settings to measure each pair of balls.

Here, the hidden variable λ represents the set of inherent properties of the balls which determine their color. For instance, it could be some internal parameters set before Alice and Bob receive them.

We assume perfect anti-correlation, meaning if one ball in the pair is white, the other is guaranteed to be black. Mathematically:

$$P(\lambda) = \begin{cases} 1, & \text{if } \lambda \Rightarrow \text{one ball white and the other black,} \\ 0, & \text{otherwise.} \end{cases} \quad (9.9)$$

The correlation functions in the Bell-CHSH quantity all equal -1 , because only the terms involving $AB = -1$ are non-zero in the following calculation:

$$E(a, b) = \sum_{\lambda} P(\lambda) A(a, \lambda) B(b, \lambda) = -1, \quad (9.10a)$$

$$E(a', b') = \sum_{\lambda} P(\lambda) A(a', \lambda) B(b', \lambda) = -1, \quad (9.10b)$$

$$E(a, b') = \sum_{\lambda} P(\lambda) A(a, \lambda) B(b', \lambda) = -1, \quad (9.10c)$$

$$E(a', b) = \sum_{\lambda} P(\lambda) A(a', \lambda) B(b, \lambda) = -1. \quad (9.10d)$$

The Bell-CHSH quantity becomes

$$S = -1 + 1 - 1 - 1 = -2. \quad (9.11)$$

This value of $|S|$ satisfies the Bell-CHSH inequality, $|S| \leq 2$.

Note that in this classical example, the outcome can be explained using local realism, where the color correlation is predetermined by some hidden variable (λ). This is in contrast to the quantum case, which we will discuss next, where the Bell-CHSH inequality can be violated, indicating the presence of nonlocal correlations or entanglement.

Exercise 9.1 Derive S for the perfectly correlated classical case, where the color of Bob's ball always matches that of Alice's for the same pair.

9.4.2 Gray Balls with Shared Ink

The Bell-CHSH inequality is not violated as long as the colors of the balls are independent of the measurement settings. However, this scenario becomes problematic when considering simultaneous measurements. If Alice and Bob make measurements at the exact same time while being separated by a distance, the pigments would not have enough time to transfer between the balls to determine their colors. This means that the outcomes are not available at the time of measurement, rendering this scenario unsuitable for testing the Bell-CHSH inequality. It fails to account for the type of correlations observed in entangled quantum systems, as it relies on a physical transfer of pigment that cannot occur instantaneously over a distance.

9.4.3 Adaptable Intelligent Balls

Consider a scenario where the balls possess robotic-like features, allowing them to sense their environment and independently choose which color to display based on the measurement settings they observe.

We still have the measurement settings (a, a', b, b') as different ways Alice and Bob decide to measure or analyze the colors of the balls (e.g., using a net, an umbrella, gloves, or bare hands.) However, in this scenario, Alice and Bob can delay their choice of measurement settings until after the balls are dispensed and just before they make the measurements.

As the balls are intelligent and can adapt independently, they observe how Alice and Bob prepare to measure them and individually set their colors to create the following outcomes:

- $E(a, b) = 1$ (Alice: bare hands; Bob: net \Rightarrow Balls: both white.)
- $E(a, b') = -1$ (Alice: bare hands; Bob: umbrella \Rightarrow Balls: black and white.)
- $E(a', b) = 1$ (Alice: gloves; Bob: net \Rightarrow Balls: both white.)
- $E(a', b') = 1$ (Alice: gloves; Bob: umbrella \Rightarrow Balls: both black.)

Now, the Bell-CHSH quantity becomes:

$$|S| = |1 - (-1) + 1 + 1| = 4. \quad (9.12)$$

As you can see, in this example, the value of $|S|$ is 4, which violates the Bell-CHSH inequality ($|S| \leq 2$). This violation occurs because the intelligent balls do not adhere to the principle of realism; their colors are not predetermined but change based on the measurement settings at the time of the measurement. Furthermore, the balls need to communicate in order to decide whether they will display identical or opposite colors, which violates the locality assumption and makes it impossible to test a pair of balls simultaneously.

This analogy helps us understand the puzzling nature of quantum entanglement, where particles seemingly violate local realism in a similar way. However, as you will see in the subsequent sections, in the quantum case, the Bell inequality is violated under simultaneous measurements, even when the particles are separated by large distances. This violation is considered evidence of the non-classical nature of quantum entanglement, which cannot be explained by any classical theory based on local hidden variables.

In the subsequent discussion on experimental verification, we will explore Bell test experiments that are ‘loophole-free’, and are specifically designed to eliminate possibilities similar to those of the adaptable intelligent balls.

9.5 Bell-CHSH Inequality with Quantum Entanglement

9.5.1 Violation of the Bell CHSH Inequality: A Special Case

In this section, we will illustrate how the Bell CHSH inequality can be violated with a specific arrangement of measurement settings for Alice and Bob using entangled qubits. A comprehensive mathematical treatment of Bell CHSH inequality for Bell states will follow in the next section.

One key feature that separates quantum mechanics from classical probability theory is the way expected values of measurements are determined in quantum mechanics – they are governed by the inner products of measurement observables.

Let’s consider two entangled qubits, A and B , in the Bell state $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. This state is often used in experiments testing Bell inequalities with photons. If Alice employs M_a to measure her qubit and Bob uses M_b , the expected value of the joint measurement, analogous to the classical counterpart in Eq. 9.5, is given by:

$$E(a, b) = \langle \Psi^- | (M_a \otimes M_b) | \Psi^- \rangle. \quad (9.13)$$

Now, let’s consider the following measurement observables for Alice and Bob:

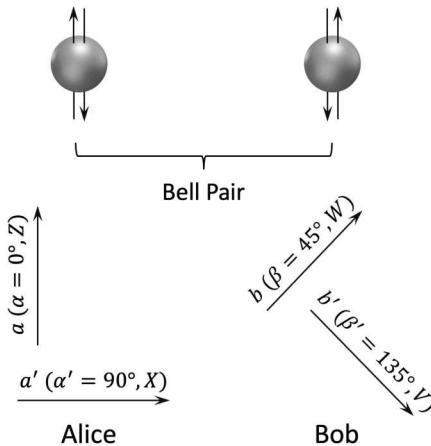


Figure 9.2: Settings for Violating the Bell-CHSH Inequality

$$M_a = Z, \quad (9.14a)$$

$$M_{a'} = X, \quad (9.14b)$$

$$M_b = W \equiv \frac{1}{\sqrt{2}}(X + Z), \quad (9.14c)$$

$$M_{b'} = V \equiv \frac{1}{\sqrt{2}}(X - Z). \quad (9.14d)$$

These measurement observables are associated with measurement angles on the Bloch sphere $\alpha = 0$, $\alpha' = \frac{\pi}{2}$, $\beta = \frac{\pi}{4}$, and $\beta' = \frac{3\pi}{4}$, as depicted in Fig. 9.2.

Observe that

$$\langle 0|Z|0\rangle = 1, \quad \langle 1|Z|1\rangle = -1, \quad \langle 0|Z|1\rangle = 0, \quad \langle 1|Z|0\rangle = 0, \quad (9.15a)$$

$$\langle 0|X|0\rangle = 0, \quad \langle 1|X|1\rangle = 0, \quad \langle 0|X|1\rangle = 1, \quad \langle 1|X|0\rangle = 1. \quad (9.15b)$$

We can then compute

$$E(a, b) = \frac{1}{\sqrt{2}}(\langle 01| - \langle 10|)Z \otimes \frac{1}{\sqrt{2}}(X + Z) \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \quad (9.16a)$$

$$= \frac{1}{2\sqrt{2}} (\langle 0|Z|0\rangle\langle 1|Z|1\rangle + \langle 1|Z|1\rangle\langle 0|Z|0\rangle) \quad (9.16b)$$

$$= -\frac{1}{\sqrt{2}}, \quad (9.16c)$$

$$E(a, b') = \frac{1}{\sqrt{2}}(\langle 01| - \langle 10|)Z \otimes \frac{1}{\sqrt{2}}(X - Z) \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \quad (9.17a)$$

$$= \frac{1}{2\sqrt{2}} (\langle 0|Z|0\rangle\langle 1|(-Z)|1\rangle + \langle 1|Z|1\rangle\langle 0|(-Z)|0\rangle) \quad (9.17b)$$

$$= \frac{1}{\sqrt{2}}, \quad (9.17c)$$

$$E(a', b) = \frac{1}{\sqrt{2}}(\langle 01| - \langle 10|)X \otimes \frac{1}{\sqrt{2}}(X + Z) \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \quad (9.18a)$$

$$= \frac{1}{2\sqrt{2}}(-\langle 0| X |1\rangle \langle 1| X |0\rangle - \langle 1| X |0\rangle \langle 0| X |1\rangle) \quad (9.18b)$$

$$= -\frac{1}{\sqrt{2}}, \quad (9.18c)$$

$$E(a', b') = \frac{1}{\sqrt{2}}(\langle 01| - \langle 10|)X \otimes \frac{1}{\sqrt{2}}(X - Z) \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \quad (9.19a)$$

$$= \frac{1}{2\sqrt{2}}(-\langle 0| X |1\rangle \langle 1| X |0\rangle - \langle 1| X |0\rangle \langle 0| X |1\rangle) \quad (9.19b)$$

$$= -\frac{1}{\sqrt{2}}. \quad (9.19c)$$

Hence, the Bell CHSH quantity is:

$$S = E(a, b) - E(a, b') + E(a', b) + E(a', b') = -2\sqrt{2}. \quad (9.20)$$

And

$$|S| > 2, \quad (9.21)$$

demonstrating that the Bell-CHSH inequality is violated. This violation is a signature of quantum entanglement and shows that local hidden-variable theories are insufficient to account for the correlations observed in quantum systems.

An Analog for the Bell Inequality Violation

Consider a triangle with sides a , b , and c . Naturally, the sum of any two sides, say $a + b$, is greater than the third side, c . However, the relationship $a^2 + b^2 > c^2$ is not universally true.

In the classical domain, if we associate probabilities with the lengths a , b , and c , the inequality $a + b > c$ represents the Bell inequality holding true.

Contrastingly, in quantum mechanics, the linearly additive quantities are probability amplitudes instead of probabilities. Thus, when probabilities are associated with a^2 , b^2 , and c^2 , the inequality $a^2 + b^2 \geq c^2$ showcases the potential violation of the Bell inequality.

Exercise 9.2 Derive the Bell CHSH quantity S for the state $|01\rangle$ with the measurement settings specified in this subsection. Note this state also gives rise to perfect anti-correlation when measured in the computational basis. Check if the Bell CHSH inequality is violated.

9.5.2 Violation of the Bell CHSH Inequality: General Case

We will defer the more general aspects of the Bell CHSH inequality and its violation in Bell states to § 9.8, given its mathematical complexities.

Key Takeaways

Bell inequalities provide a mathematical framework that allows for experiments to differentiate between these two cases: hidden variable theories and quantum theory of entanglement. These inequalities establish limits on the correlations that can be observed if the entangled particles followed local hidden variable theories. Experiments testing Bell inequalities have consistently shown that the predictions of quantum mechanics are correct, and the correlations between entangled particles are better explained by the non-local, instantaneous nature of quantum mechanics rather than local hidden variable theories.

9.6 Experimental Verification

Experimental verification of Bell's inequality, also known as Bell tests, refers to tests of local realism against the predictions of quantum mechanics using entangled particles. These experiments are designed to measure the correlations between the outcomes of spatially separated particles to check whether they obey the constraints set by Bell inequalities, such as the Bell-CHSH inequality. If the inequalities are violated, it implies that local realism is incompatible with the observed behavior, and the particles exhibit nonlocal correlations characteristic of quantum entanglement.

9.6.1 *Basic Experimental Setup

A basic experimental setup to test the EPR paradox is illustrated in Fig. 9.3. The entangled photon pairs are generated through spontaneous parametric down-conversion (SPDC), which is a widely-used method to produce pairs of entangled photons for various quantum communication and computation experiments. In SPDC, a single high-energy photon (labeled as pump laser) passes through a nonlinear BBO crystal, and occasionally, it splits into two lower-energy photons that are entangled. The entangled photons can be created in various Bell states, depending on the experiment setup and the properties of the nonlinear crystal. The state is commonly $\Psi^- = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, which means in terms of photon polarization:

$$\Psi^- = \frac{1}{\sqrt{2}}(|\leftrightarrow\rangle_1 \otimes |\downarrow\rangle_2 - |\uparrow\rangle_1 \otimes |\leftrightarrow\rangle_2), \quad (9.22)$$

where $|\leftrightarrow\rangle_1$, $|\uparrow\rangle_1$, $|\leftrightarrow\rangle_2$, and $|\downarrow\rangle_2$ represent the horizontal and vertical polarizations of the two photons in the pair.

The two photons in each pair are sent to Alice and Bob, respectively, who are located at spatially separated locations. The transmission media can be free space or optical fiber, spanning distances up to hundreds of kilometers.

Both Alice and Bob pass their respective photons through a beam splitter (labeled BS in Fig. 9.3), which is essentially a half mirror. The beam splitter sends the photon in two different paths, corresponding to measurement bases a and a' for Alice, and b and b' for Bob, randomly. The splitter can also be replaced by an optical switch, which can be controlled by Alice or Bob to select the measurement bases.

The angles of the measurement bases (α, α' for Alice and β, β' for Bob) are adjusted using half-wave plates (HWP). The photon then arrives at a polarization beam splitter (PBS). The PBS sends the photon to different detectors according to

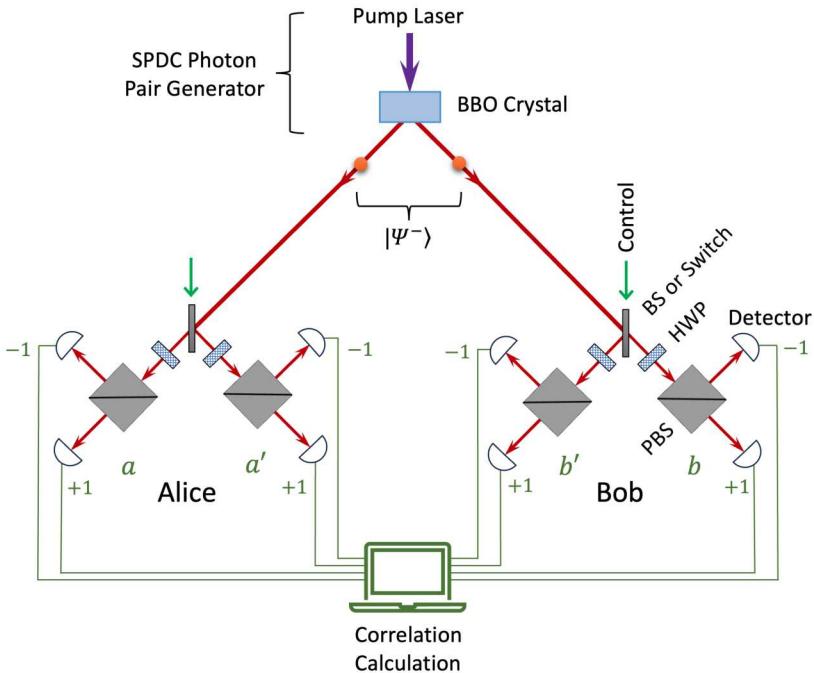


Figure 9.3: Basic Experimental Setup for Bell Test

its polarization. The measurement results (always $+1$ or -1) are fed into a computer to analyze the correlations and calculate the Bell-CHSH quantity.

The experimental setup described above allows testing the Bell-CHSH inequality by measuring correlations between the polarization states of entangled photon pairs. By varying the settings of the measurement bases, one can obtain different values of the Bell-CHSH quantity S and check for violation of the Bell-CHSH inequality.

9.6.2 Early Experiments

Early experiments aimed at verifying Bell's inequality played a crucial role in establishing the empirical basis for quantum entanglement. Notable among these are:

Aspect experiments (1981-1982) [16], [17]: Alain Aspect and his team conducted a series of groundbreaking experiments using entangled photons to test the Bell-CHSH inequality. They employed the technique of time-coincidence measurements, which involves ensuring that detections are only counted when they could be paired as originating from the same event, and observed strong correlations between the polarization states of the photons, which violated the Bell-CHSH inequality. These results provided strong evidence in favor of quantum mechanics over local hidden variable theories.

Weihs et al. (1998) [89]: In this experiment, Gregor Weihs and his team tested the Bell-CHSH inequality with entangled photon pairs over a distance of about 400 meters. They used a high-efficiency source of entangled photon pairs and fast-

switching polarization analyzers. Their results showed a significant violation of the Bell-CHSH inequality, ruling out local hidden variable theories even for particles separated by large distances.

9.6.3 Loophole-free Measurements

There are several loopholes, or measurement flaws, that have been identified in earlier Bell test experiments:

Detection (or Fair Sampling) loophole: This loophole arises when the detectors used in the experiment have low efficiency or when there is a bias in the detection process. It suggests that the observed correlations could be a result of selective sampling rather than genuine quantum entanglement.

Communication (or Locality) loophole: This loophole arises when there is insufficient space or time separation between the measurement events, allowing for the possibility of classical communication between the detectors or the particles during the experiment.

Freedom-of-choice loophole: This loophole arises when the choice of measurement settings is not truly random or when there is a possibility that the measurement settings could be influenced by some unknown variables correlated with the particles.

“Loophole-free” Bell test experiments are designed to test the predictions of quantum mechanics against local hidden variable theories without being affected by any of the known loopholes.

Hensen et al. (2015) [54]: This experiment, also known as the “loophole-free Bell test,” addressed some of the potential loopholes in previous experiments, such as the detection loophole and the locality loophole. Unlike the previous experiments that used entangled photons, Hensen and his team used entangled electron spins in diamond nitrogen-vacancy centers separated by 1.3 kilometers, which offered different experimental advantages. They observed a statistically significant violation of the Bell-CHSH inequality, providing strong evidence against local realism.

Giustina et al. (2015) [45] and Shalm et al. (2015) [80]: These two experiments, conducted independently by different research groups, addressed the detection loophole and the locality loophole simultaneously. Both experiments used entangled photons and highly efficient detectors, ensuring that the observed correlations could not be explained by local hidden variable theories. The results of both experiments showed a clear violation of the Bell-CHSH inequality.

By separating the entangled particles by a significant distance and switching the measurement settings quickly, these experiments aimed to rule out the possibility of any local interaction or communication between the particles that could influence their correlated outcomes. This addresses the locality loophole, which suggests that entangled particles might exchange information or be influenced by a common cause in their local environment, thereby explaining the observed correlations.

9.6.4 Key Experimental Findings and Implications

The loophole-free experiments offer compelling evidence for the nonlocal character of quantum entanglement, a fundamental aspect in the comprehension of quantum

phenomena. The consistent experimental confirmations of Bell's inequalities are an important cornerstone in the development of quantum mechanics.

These experiments have profound implications for our understanding of the fundamental nature of the universe. They challenge the premise of local realism, which posits that physical properties are predetermined by the local environment and unaffected by measurement, and that no physical influence can travel faster than light. In contrast, the experimental results align with quantum mechanics, which predicts nonlocal correlations and intrinsic randomness.

Rejecting local realism leads to two possibilities: either accepting the concept of superluminal communication, akin to speculative notions found in science fiction but lacking scientific basis, or embracing the idea of quantum entanglement being inherently nonlocal. The latter is widely accepted in the scientific community, reinforcing the paradigm shift in our understanding of quantum interactions and their implications for the nature of reality.

Furthermore, the intrinsic randomness observed in quantum measurements, as quantified by Born's rule (§ 1.5.1), underscores a departure from classical determinism. Born's rule, which assigns probabilities to the outcomes of quantum measurements, is not a reflection of our ignorance about the system's state but a fundamental aspect of quantum mechanics. This probabilistic nature reveals that, unlike classical systems where uncertainty often stems from incomplete knowledge, quantum randomness is an inherent feature of the universe.

9.6.5 2022 Nobel Prize in Physics

The Nobel Prize in Physics 2022 was awarded jointly to Alain Aspect, John F. Clauser, and Anton Zeilinger for experiments with entangled photons, establishing the violation of Bell inequalities and pioneering quantum information science.

Alain Aspect, John Clauser, and Anton Zeilinger are three prominent physicists who have made significant contributions to the experimental verification of quantum entanglement and the testing of local realism through Bell inequalities. Their work earned them the Nobel Prize in Physics in 2022. Aspect's groundbreaking experiments in the 1980s provided strong evidence in favor of quantum mechanics over local hidden variable theories. Clauser, along with Stuart Freedman, performed one of the first experimental tests of Bell's inequality in the early 1970s, which showed results consistent with the predictions of quantum mechanics. Zeilinger has been a pioneer in the field of quantum information and has conducted numerous experiments on entanglement, pushing the boundaries of our understanding of quantum mechanics. The groundbreaking work of these physicists has been instrumental in shaping our current understanding of the nonlocal nature of quantum entanglement and the fundamental nature of the universe.

As highlighted by the 2022 Nobel Prize, the experimental verification of Bell's inequalities has been central to the development and confirmation of quantum theory. It has challenged our classical intuitions and has opened new avenues for understanding the fundamental aspects of the physical world. As our understanding and technological capabilities evolve, future experiments may offer new insights into the nature of quantum entanglement and its implications for our understanding of the physical world. Additionally, the practical applications of entanglement in the fields of quantum computing, quantum communication, and quantum cryptography

make this area of research not only fundamentally interesting but also highly relevant to emerging technologies.

9.7 The No-Communication Theorem

Quantum entanglement and the violation of Bell's inequalities raise questions regarding the possibility of transmitting information faster than the speed of light, a concept often dubbed “superluminal communication.” While entangled particles exhibit nonlocal correlations, it's essential to recognize that these correlations do not allow for instantaneous information transfer between distant observers. Nonlocal correlations cannot be exploited for superluminal communication, preserving the fundamental principle that information cannot travel faster than light, and thereby maintaining the core tenets of special relativity and causality. This is formally stated as the no-communication theorem which is a key result in quantum information theory:

The No-Communication Theorem.

During the measurement of an entangled quantum state, it is fundamentally impossible for one observer to communicate information to another solely through the act of measurement.

Let's consider a scenario where two observers, Alice and Bob, each hold one qubit of a two-qubit entangled state. If Alice measures her qubit, the state of her qubit collapses to a specific value. However, this does not immediately change the state of Bob's qubit in any way that Bob can observe.

The no-communication theorem, despite the nonlocality of quantum mechanics, holds true primarily because the measurement outcome in quantum mechanics is intrinsically random. While Alice's measurement collapses the overall state from her perspective, Bob, without additional information, can only see a random outcome when he measures his qubit. He cannot distinguish whether the randomness is due to quantum uncertainty or because of Alice's prior measurement.

To illustrate this further, consider the Ekert 91 (E91) quantum key distribution (QKD) protocol (see § 10.6 for details). In this protocol, Alice and Bob use entangled pairs of qubits to generate a shared secret key. At first glance, it might seem as if Alice can instantaneously communicate secret key bits to Bob, especially if they have previously agreed on a sequence of measurement bases. However, the key they generate is a string of random bits, which carries no information about Alice's state. It's only after Alice communicates with Bob through a classical channel (which cannot exceed the speed of light) to discuss the measurement results that the key becomes useful.

A proof of the No-Communication Theorem is given in § 12.2.9.

9.7.1 Myths and Misconceptions

It is important to note that quantum entanglement is a well-defined concept within the realm of quantum mechanics, and it should not be conflated with unrelated phenomena. There are several myths and misconceptions surrounding quantum entanglement, often stemming from a misunderstanding of the underlying principles

of quantum mechanics or from misapplying quantum phenomena to everyday life. Here are a few examples:

- Instantaneous communication through entanglement: Although entanglement can create strong correlations between particles regardless of the distance between them, it cannot be used to transmit information instantaneously, as this would violate the principle of causality and the speed of light limit imposed by special relativity. The process of measurement and the collapse of entangled states still follow the constraints of relativistic causality.
- Macroscopic objects can be easily entangled: Entanglement is typically observed at the level of subatomic particles, and it becomes increasingly difficult to maintain and observe entanglement as the size of the system increases. While there have been some experimental demonstrations of entanglement in larger systems, entangling macroscopic objects like everyday items or living organisms is extremely challenging due to decoherence and environmental factors.
- Entanglement as a source of paranormal phenomena: Some people may claim that quantum entanglement is responsible for various paranormal phenomena, such as telepathy or psychic powers. There are also claims of quantum entanglement between people and their spirit, between dreams and reality, and so on. However, there is no scientific evidence to support such claims. Entanglement is a well-established concept in quantum mechanics, but it has not been shown to have any connection to paranormal phenomena.

9.8 *Derivation of Bell-CHSH Quantity for Bell States

In this section, we present a more general, step by step analysis of the Bell CHSH inequality. We use the Bell state $|\Psi^-\rangle$ for this purpose; other Bell states follow similarly.

9.8.1 *Derivation via Measurement Bases

1 Measurement Bases and Observables

Suppose a qubit is measured along a direction defined by the polar angle α and the azimuthal angle θ on the Bloch sphere. Effectively, it is measured using the basis $\{|a\rangle, |a_\perp\rangle\}$:

$$|a\rangle = \cos \frac{\alpha}{2} |0\rangle + \sin \frac{\alpha}{2} |1\rangle, \quad (9.23a)$$

$$|a_\perp\rangle = -\sin \frac{\alpha}{2} |0\rangle + \cos \frac{\alpha}{2} |1\rangle. \quad (9.23b)$$

The associated measurement observables can be represented as:

$$M_a = |a\rangle\langle a| - |a_\perp\rangle\langle a_\perp|. \quad (9.24)$$

In the Bell CHSH inequality, there are four bases and four measurement observables involved, similar to the above. For each pair of qubits, Alice selects one of the two bases, $\{|a\rangle, |a_\perp\rangle\}$ or $\{|a'\rangle, |a'_\perp\rangle\}$, to measure qubit A:

$$|a\rangle = \cos \frac{\alpha}{2} |0\rangle + \sin \frac{\alpha}{2} |1\rangle, \quad |a_{\perp}\rangle = -\sin \frac{\alpha}{2} |0\rangle + \cos \frac{\alpha}{2} |1\rangle, \quad (9.25a)$$

$$|a'\rangle = \cos \frac{\alpha'}{2} |0\rangle + \sin \frac{\alpha'}{2} |1\rangle, \quad |a'_{\perp}\rangle = -\sin \frac{\alpha'}{2} |0\rangle + \cos \frac{\alpha'}{2} |1\rangle. \quad (9.25b)$$

Similarly, Bob chooses between the bases $\{|b\rangle, |b_{\perp}\rangle\}$ and $\{|b'\rangle, |b'_{\perp}\rangle\}$ for measuring qubit B:

$$|b\rangle = \cos \frac{\beta}{2} |0\rangle + \sin \frac{\beta}{2} |1\rangle, \quad |b_{\perp}\rangle = -\sin \frac{\beta}{2} |0\rangle + \cos \frac{\beta}{2} |1\rangle, \quad (9.26a)$$

$$|b'\rangle = \cos \frac{\beta'}{2} |0\rangle + \sin \frac{\beta'}{2} |1\rangle, \quad |b'_{\perp}\rangle = -\sin \frac{\beta'}{2} |0\rangle + \cos \frac{\beta'}{2} |1\rangle. \quad (9.26b)$$

These four measurement bases correspond to the angles α, α', β , and β' on the Bloch sphere.

The corresponding measurement observables are:

$$M_a = |a\rangle\langle a| - |a_{\perp}\rangle\langle a_{\perp}|, \quad (9.27a)$$

$$M_{a'} = |a'\rangle\langle a'| - |a'_{\perp}\rangle\langle a'_{\perp}|, \quad (9.27b)$$

$$M_b = |b\rangle\langle b| - |b_{\perp}\rangle\langle b_{\perp}|, \quad (9.27c)$$

$$M_{b'} = |b'\rangle\langle b'| - |b'_{\perp}\rangle\langle b'_{\perp}|. \quad (9.27d)$$

When Alice measures qubit A and Bob measures qubit B, they employ joint measurement observables like the following:

$$M_{ab} = M_a \otimes M_b. \quad (9.28)$$

2 Measurement Expected Values

To check the Bell-CHSH inequality, we will calculate the expected values of the products of the outcomes, or correlation functions, $E(a, b)$, $E(a, b')$, $E(a', b)$, and $E(a', b')$. Note in quantum mechanics, expected values are computed using inner products, which is a fundamental distinction from classical mechanics:

$$E(a, b) = \langle \Psi^- | (M_a \otimes M_b) | \Psi^- \rangle, \quad (9.29a)$$

$$E(a, b') = \langle \Psi^- | (M_a \otimes M_{b'}) | \Psi^- \rangle, \quad (9.29b)$$

$$E(a', b) = \langle \Psi^- | (M_{a'} \otimes M_b) | \Psi^- \rangle, \quad (9.29c)$$

$$E(a', b') = \langle \Psi^- | (M_{a'} \otimes M_{b'}) | \Psi^- \rangle. \quad (9.29d)$$

After some algebraic manipulations (see § 9.8.1.5), we obtain:

$$E(a, b) = -\cos(\alpha - \beta), \quad (9.30a)$$

$$E(a, b') = -\cos(\alpha - \beta'), \quad (9.30b)$$

$$E(a', b) = -\cos(\alpha' - \beta), \quad (9.30c)$$

$$E(a', b') = -\cos(\alpha' - \beta'). \quad (9.30d)$$



The result in Eq. 9.30 is remarkable in its own right. It implies that the measurement correlations solely depend on the difference between the measurement angles,

rather than their individual values. For example, if $\alpha = \beta$, then $E(a, b) = -1$, signifying that the state $|\Psi^-\rangle$ exhibits isotropic anti-correlations. This is further supported by the equation:

$$\begin{aligned} |\Psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \\ &= \frac{1}{\sqrt{2}}(|aa_\perp\rangle - |a_\perp a\rangle). \end{aligned}$$

This is quite different from correlated classical spins. In fact, this is an underlying reason for the violation of the Bell inequality.

3 The Bell CHSH Quantity

Inserting the expressions from Eq. 9.30 into the Bell-CHSH quantity in Eq. 9.8 we obtain:

$$S = E(a, b) - E(a, b') + E(a', b) + E(a', b') \quad (9.31a)$$

$$= -\cos(\alpha - \beta) + \cos(\alpha - \beta') - \cos(\alpha' - \beta) - \cos(\alpha' - \beta'). \quad (9.31b)$$

Exercise 9.3 Derive expressions of the Bell-CHSH quantity S (similar to Eq. 9.31) for the other three Bell states.

4 Violation of the Bell CHSH Inequality

Upon examining Eq. 9.31, we observe that it is feasible to select angles α , α' , β , and β' such that $|S|$ exceeds 2, thereby violating the Bell-CHSH inequality. Specifically, by setting $\alpha = 0$, $\beta = \frac{\pi}{4}$, $\alpha' = \frac{\pi}{2}$, and $\beta' = \frac{3\pi}{4}$ (as depicted in Fig. 9.2), we find that:

$$|S| = \frac{1}{2} \left| \sqrt{2} - (-\sqrt{2}) + \sqrt{2} + \sqrt{2} \right| = 2\sqrt{2} > 2. \quad (9.32)$$

This result illustrates that the Bell-CHSH inequality can be violated by entangled qubits in a quantum scenario, which implies that local hidden-variable theories are insufficient in fully accounting for the correlations in quantum systems.



The measurement observables M_a , M'_a , M_b , and M'_b can be expressed as linear combinations of Pauli operators. For the angles specified above, they correspond to Pauli operators or their simple linear combinations given by Eq. 9.14.

Exercise 9.4 Consider the product state

$$|++\rangle \equiv |+\rangle \otimes |+\rangle = \frac{1}{2}(|00\rangle + |11\rangle + |01\rangle + |10\rangle).$$

Derive S for this state as a function of α and β . Investigate if the Bell CHSH inequality can be violated with this state.

5 Derivation of Eq. 9.30

Start with the expected values of M_a across the one-qubit basis states:

$$\langle 0 | M_a | 0 \rangle = \cos \alpha, \quad (9.33a)$$

$$\langle 1 | M_a | 1 \rangle = -\cos \alpha, \quad (9.33b)$$

$$\langle 1 | M_a | 0 \rangle = \sin \alpha, \quad (9.33c)$$

$$\langle 0 | M_a | 1 \rangle = \sin \alpha. \quad (9.33d)$$

Now calculate the expected values of $M_a \otimes M_b$ across the two-qubit basis states:

$$\langle 11 | (M_a \otimes M_b) | 11 \rangle = \langle 1 | M_a | 1 \rangle \langle 1 | M_b | 1 \rangle = \cos \alpha \cos \beta, \quad (9.34a)$$

$$\langle 00 | (M_a \otimes M_b) | 00 \rangle = \langle 0 | M_a | 0 \rangle \langle 0 | M_b | 0 \rangle = \cos \alpha \cos \beta, \quad (9.34b)$$

$$\langle 11 | (M_a \otimes M_b) | 00 \rangle = \langle 1 | M_a | 0 \rangle \langle 1 | M_b | 0 \rangle = \sin \alpha \sin \beta, \quad (9.34c)$$

$$\langle 00 | (M_a \otimes M_b) | 11 \rangle = \langle 0 | M_a | 1 \rangle \langle 0 | M_b | 1 \rangle = \sin \alpha \sin \beta, \quad (9.34d)$$

$$\langle 01 | (M_a \otimes M_b) | 01 \rangle = \langle 0 | M_a | 0 \rangle \langle 1 | M_b | 1 \rangle = -\cos \alpha \cos \beta, \quad (9.34e)$$

$$\langle 10 | (M_a \otimes M_b) | 10 \rangle = \langle 1 | M_a | 1 \rangle \langle 0 | M_b | 0 \rangle = -\cos \alpha \cos \beta, \quad (9.34f)$$

$$\langle 01 | (M_a \otimes M_b) | 10 \rangle = \langle 0 | M_a | 1 \rangle \langle 1 | M_b | 0 \rangle = \sin \alpha \sin \beta, \quad (9.34g)$$

$$\langle 10 | (M_a \otimes M_b) | 01 \rangle = \langle 1 | M_a | 0 \rangle \langle 0 | M_b | 1 \rangle = \sin \alpha \sin \beta. \quad (9.34h)$$

Finally, calculate the expected values of $M_a \otimes M_b$ for the Bell states:

$$\langle \Phi^+ | (M_a \otimes M_b) | \Phi^+ \rangle = \cos \alpha \cos \beta + \sin \alpha \sin \beta = \cos(\alpha - \beta), \quad (9.35a)$$

$$\langle \Phi^- | (M_a \otimes M_b) | \Phi^- \rangle = \cos \alpha \cos \beta - \sin \alpha \sin \beta = \cos(\alpha + \beta), \quad (9.35b)$$

$$\langle \Psi^+ | (M_a \otimes M_b) | \Psi^+ \rangle = -\cos \alpha \cos \beta + \sin \alpha \sin \beta = -\cos(\alpha + \beta), \quad (9.35c)$$

$$\langle \Psi^- | (M_a \otimes M_b) | \Psi^- \rangle = -\cos \alpha \cos \beta - \sin \alpha \sin \beta = -\cos(\alpha - \beta). \quad (9.35d)$$

9.8.2 *Derivation Using Pauli Operators

This subsection outlines a general derivation of the Bell CHSH inequality using Pauli matrices (X, Y, Z). This approach is based on expressing measurement operators as linear combinations of the Pauli matrices. Even though more abstract, this approach is a powerful technique that can be used to analyze more general scenarios beyond the Bell inequalities.

1 Outline of the Derivation

The measurement operator in the direction specified by a 3D unit vector $\mathbf{a} = (a_1, a_2, a_3)$ is given by:

$$M_{\mathbf{a}} = \mathbf{a} \cdot \boldsymbol{\sigma} \quad (9.36a)$$

$$\equiv a_1 \sigma_1 + a_2 \sigma_2 + a_3 \sigma_3, \quad (9.36b)$$

where $\sigma_1 \equiv X$, $\sigma_2 \equiv Y$, and $\sigma_3 \equiv Z$ are the Pauli operators.

Suppose Alice measures her qubit along directions \mathbf{a} and \mathbf{a}' , and Bob \mathbf{b} and \mathbf{b}' . Then the measurement correlation functions are given by:

$$E(\mathbf{a}, \mathbf{b}) = \langle \Psi^- | (M_{\mathbf{a}} \otimes M_{\mathbf{b}}) | \Psi^- \rangle, \quad (9.37a)$$

$$E(\mathbf{a}, \mathbf{b}') = \langle \Psi^- | (M_{\mathbf{a}} \otimes M_{\mathbf{b}'}) | \Psi^- \rangle, \quad (9.37b)$$

$$E(\mathbf{a}', \mathbf{b}) = \langle \Psi^- | (M_{\mathbf{a}'} \otimes M_{\mathbf{b}}) | \Psi^- \rangle, \quad (9.37c)$$

$$E(\mathbf{a}', \mathbf{b}') = \langle \Psi^- | (M_{\mathbf{a}'} \otimes M_{\mathbf{b}'}) | \Psi^- \rangle. \quad (9.37d)$$

With much algebra (see subsection below), it can be shown:

$$E(\mathbf{a}, \mathbf{b}) = -\mathbf{a} \cdot \mathbf{b} \quad (9.38a)$$

$$= -(a_1 b_1 + a_2 b_2 + a_3 b_3) \quad (9.38b)$$

$$= -\cos \phi_{\mathbf{ab}}, \quad (9.38c)$$

where $\phi_{\mathbf{ab}}$ represents the angle between \mathbf{a} and \mathbf{b} .

And similarly for the other correlation functions. These results mirror Eq. 9.30.

The Bell CHSH quantity is:

$$S = E(\mathbf{a}, \mathbf{b}) - E(\mathbf{a}, \mathbf{b}') + E(\mathbf{a}', \mathbf{b}) + E(\mathbf{a}', \mathbf{b}') \quad (9.39a)$$

$$= -\cos \phi_{\mathbf{ab}} + \cos \phi_{\mathbf{ab}'} - \cos \phi_{\mathbf{a}'\mathbf{b}} - \cos \phi_{\mathbf{a}'\mathbf{b}'} \quad (9.39b)$$

Now we can pick non-coplanar measurement directions to demonstrate the violation of Bell CHSH inequality. A particular choice is:

$$\mathbf{a} = (1, 0, 0), \quad (9.40a)$$

$$\mathbf{a}' = (0, 1, 0), \quad (9.40b)$$

$$\mathbf{b} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right), \quad (9.40c)$$

$$\mathbf{b}' = \left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right). \quad (9.40d)$$

This yields $S = \frac{4}{\sqrt{3}}$ which is smaller than $2\sqrt{2}$ but still greater than 2.

2 Derivation of Eq. 9.38

Let us consider the correlation function $E(\mathbf{a}, \mathbf{b})$:

$$E(\mathbf{a}, \mathbf{b}) = \langle \Psi^- | (M_{\mathbf{a}} \otimes M_{\mathbf{b}}) | \Psi^- \rangle \quad (9.41a)$$

$$= \langle \Psi^- | ((a_1 \sigma_1 + a_2 \sigma_2 + a_3 \sigma_3) \otimes (b_1 \sigma_1 + b_2 \sigma_2 + b_3 \sigma_3)) | \Psi^- \rangle \quad (9.41b)$$

$$= \sum_{i=1}^3 \sum_{j=1}^3 a_i b_j \langle \Psi^- | (\sigma_i \otimes \sigma_j) | \Psi^- \rangle. \quad (9.41c)$$

Now, it can be shown that

$$\langle \Psi^- | (\sigma_i \otimes \sigma_j) | \Psi^- \rangle = -\delta_{ij}. \quad (9.42)$$

While we will not delve into the detailed algebra required to prove this equation here, it is important to note the following insight. This equation indicates that there is a perfect anti-correlation in the Bell state $|\Psi^-\rangle$ when Alice and Bob measure their qubits along the same direction, meaning they will always obtain opposite outcomes. However, if the measurement directions are orthogonal, the correlation function is zero.

Substituting this result back into Eq. 9.41, we obtain:

$$E(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^3 a_i b_i \langle \Psi^- | (\sigma_i \otimes \sigma_i) | \Psi^- \rangle \quad (9.43a)$$

$$= -a_1 b_1 - a_2 b_2 - a_3 b_3, \quad (9.43b)$$

which matches Eq. 9.38.

We can derive $E(\mathbf{a}, \mathbf{b})$ for other Bell states similarly. Here is a summary:

$$\langle \Phi^+ | (M_a \otimes M_b) | \Phi^+ \rangle = a_1 b_1 - a_2 b_2 + a_3 b_3, \quad (9.44a)$$

$$\langle \Psi^+ | (M_a \otimes M_b) | \Psi^+ \rangle = a_1 b_1 + a_2 b_2 - a_3 b_3, \quad (9.44b)$$

$$\langle \Phi^- | (M_a \otimes M_b) | \Phi^- \rangle = -a_1 b_1 + a_2 b_2 + a_3 b_3, \quad (9.44c)$$

$$\langle \Psi^- | (M_a \otimes M_b) | \Psi^- \rangle = -a_1 b_1 - a_2 b_2 - a_3 b_3. \quad (9.44d)$$

9.9 * Further Exploration

For readers interested in deepening their understanding of entanglement and Bell inequalities, the following research papers and books are suggested:

1. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? [38]: In this famous paper, Einstein, Podolsky, and Rosen question the completeness of quantum mechanics and introduce the concept of “elements of reality”.
2. On the Einstein Podolsky Rosen Paradox [20]: This is John Bell’s seminal paper where he introduces Bell’s inequalities and provides a way to test the validity of local hidden variable theories against quantum mechanics.
3. Quantum Entanglement and Bell Inequalities [72]: This book chapter by Asher Peres provides an in-depth treatment of quantum entanglement and Bell inequalities, along with various proofs and experimental considerations.
4. Foundations and Applications of Quantum Entanglement [55]: This review by Horodecki et al. provides a comprehensive overview of the theoretical foundations and various applications of quantum entanglement. The authors cover entanglement measures, entanglement transformations, and the use of entanglement in quantum information processing.
5. Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment [15]: This landmark paper by Alain Aspect and colleagues reports the experimental realization of the Einstein-Podolsky-Rosen-Bohm gedankenexperiment and the observation of Bell’s inequality violation with high statistical significance.
6. Loophole-free Bell inequality violation using electron spins [54]: In this paper, Hensen et al. present a loophole-free Bell test that addresses the detection and locality loopholes by using entangled electron spins in diamond nitrogen-vacancy centers.
7. Loopholes in Bell Inequality Tests of Local Realism [65]: This review by Larsson discusses the various loopholes that can occur in experimental tests of Bell inequalities and how they can be addressed. The article provides insights into the challenges of performing loophole-free Bell tests.

8. Challenging Local Realism with Human Choices [53]: In this paper, Handsteiner et al. present an experiment where human choices are used to set the measurement settings in a Bell test. This experiment addresses the freedom-of-choice loophole, making it one of the most stringent tests of local realism.

9.10 Summary and Conclusions

The Enigma of Quantum Entanglement

The chapter set forth a profound exploration into the realm of quantum entanglement, an inherently quantum phenomenon that sets quantum systems apart from their classical counterparts. The inseparable correlation exhibited by entangled particles challenges classical notions, particularly local realism, leading to far-reaching implications in both the theoretical and experimental domains of quantum mechanics. Through an examination of the Einstein-Podolsky-Rosen (EPR) paradox, the chapter underscored the tension between quantum mechanics and classically intuitive perspectives on reality.

Bell Inequalities: Bridging Theory and Experiment

Central to the discourse on quantum entanglement are the Bell inequalities, which serve as invaluable mathematical tools to distinguish quantum correlations from classical ones. Specifically, the Bell-CHSH inequality, a particular form of Bell inequalities, emerged as a focal point in our discussions. Experiments testing these inequalities have consistently highlighted the nonlocal correlations between entangled particles, emphasizing that these correlations are intrinsically quantum and cannot be explained by local hidden variable theories. The reaffirmation of the nonlocal nature of quantum entanglement by loophole-free measurements has further solidified the understanding that quantum mechanics provides a more accurate description of reality than classical intuitions might suggest.

The No-Communication Theorem

While quantum entanglement might seem like a conduit for faster-than-light communication given its nonlocal characteristics, the no-communication theorem elucidates that this is not the case. This theorem effectively ensures that quantum mechanics remains consistent with relativity, barring the instantaneous transmission of information through quantum measurements.

A Methodical Dive into Bell-CHSH Inequality

For those academically inclined, the chapter probed deeper into the mathematical intricacies of the Bell-CHSH inequality using Bell states, exemplifying the conceptual richness and rigor of the subject.

Upcoming Topics

As we progress into the next chapter, our emphasis will shift from the theoretical underpinnings of quantum entanglement to its versatile applications. We will journey through key concepts like superdense coding and quantum teleportation, unravelling how quantum entanglement can be harnessed to facilitate groundbreaking advancements in quantum communication and computation. The indispensable role of quantum teleportation, in particular, will be accentuated, given its foundational

importance in the realm of quantum technologies.

This transition from the foundational understanding of entanglement to its practical applications in the next chapter will provide a holistic view, bridging the gap between abstract quantum principles and their real-world implications.

Problem Set 9

9.1 Investigate the following questions and elaborate your answers:

- (a) Do entangled particles affect each other instantaneously over vast distances?
- (b) How do “hidden variables” relate to Bell inequalities?
- (c) What is the significance of “loophole-free” Bell tests?
- (d) How do the findings of Bell-test experiments relate to the no-superluminal communication constraint?
- (e) Have there been experiments confirming the violation of Bell inequalities using qubits other than photons?
- (f) Are there practical applications for the violation of Bell inequalities, or is the phenomenon only of theoretical interest?
- (g) The violation of Bell inequalities suggests that either realism or locality is not applicable in the quantum realm. Which one is it, or are both concepts challenged?
- (h) Are there any interpretations of quantum mechanics that are consistent with both Bell inequality violations and a form of realism?
- (i) Is it possible to have a local but non-realistic theory that accurately describes quantum mechanics?
- (j) Does assigning a probability to a measurement outcome, as per Born’s rule, indicate a lack of information about the system, or are quantum measurements intrinsically random?
- (k) If Bob holds a qubit from a Bell pair, and Alice holds the other, and they are separated by vast distances, say Mars and Jupiter, will Alice be able to detect if Bob measures his qubit?
- (l) How does quantum entanglement contribute to quantum computing and communication?

9.2 Show that, for the Bell state $|\Phi^-\rangle$, the maximum of $|S|$ is $2\sqrt{2}$. Find all combinations of α , α' , β , and β' such that $|S| = 2\sqrt{2}$.

9.3 Consider the product state

$$|++\rangle \equiv |+\rangle \otimes |+\rangle = \frac{1}{2}(|00\rangle + |11\rangle + |01\rangle + |10\rangle),$$

where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Derive the Bell-CHSH quantity S (Eq. 9.8) for this state using the measurement settings specified in Eq. 9.14. Check if the Bell CHSH inequality is violated.

9.4 The CHSH Game is a thought experiment involving two players, Alice and Bob, who are separated by a great distance and cannot communicate with each other. The game is played with a referee who provides inputs and determines the outcome.

Rules of the Game:

- Inputs: The referee randomly gives Alice and Bob each a single bit (0 or 1) as input.
- Outputs: Alice and Bob each provide a single bit (0 or 1) as output, without knowing the other player's input.
- Winning Condition: Alice and Bob win the round if the following rule is satisfied: The XOR of their outputs must equal the AND of their inputs. In other words, their outputs should be different (01 or 10) only when both received a 1 for inputs.

Strategies and Winning Probabilities:

- Classical Strategy (Local Hidden Variables): In the classical world, Alice and Bob could pre-agree on a strategy or have some locally shared information ("hidden variables") to coordinate their answers.
- Quantum Strategy (Entanglement): Alice and Bob share an entangled pair of qubits and perform specific measurements based on their inputs.

Your tasks are to demonstrate:

- (a) The best classical strategy leads to a maximum winning probability of 75%.
- (b) With the quantum strategy, they can achieve a winning probability of approximately 85%!

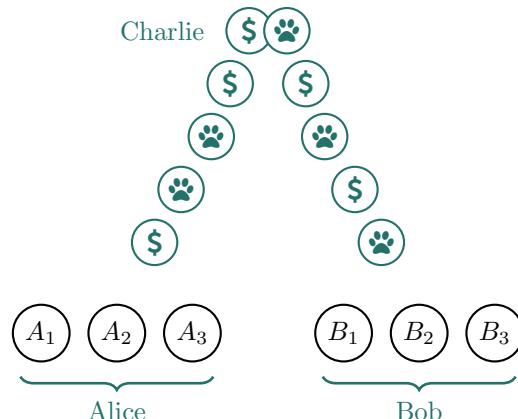


Figure 9.4: Illustration for Problem 9.5

9.5 Three-Qubit Bell Inequality. Charlie prepares many pairs of coins and sends one coin of each pair to Alice and the other to Bob. These coins are set to land randomly as Heads (H) or Tails (T), each maintaining perfect correlation or anticorrelation with its paired coin.

Alice and Bob each have three landing pads for the coins, labeled A_1, A_2, A_3 for Alice and B_1, B_2, B_3 for Bob, respectively. Each coin lands on a pad matched with the pad of its pair of the other participant (i.e., a coin on A_2 is matched with the coin on B_2), revealing either Heads (H) or Tails (T).

Alice and Bob verify the perfect correlation of their coin pairs by confirming that:

$$P(A_1 = B_1) = P(A_2 = B_2) = P(A_3 = B_3) = 1,$$

and perfect anticorrelation by confirming that:

$$P(A_1 \neq B_1) = P(A_2 \neq B_2) = P(A_3 \neq B_3) = 1.$$

They also compute a correlation C from many observations:

$$C = P(A_1 = B_2) + P(A_2 = B_3) + P(A_3 = B_1).$$

Here, $P(A_i = B_j)$ denotes the probability that Alice's coin on pad A_i and Bob's coin on pad B_j show the same face, i.e., both are Heads (H, H) or both are Tails (T, T). Show that:

- (a) If different pairs of coins are also perfectly correlated, then $C = 3$.
- (b) If the coins alternate between being perfectly correlated and perfectly anticorrelated across the pairs, then $C = 1$.
- (c) In the general case, C satisfies $C \geq 1$.

Here, $C \geq 1$ represents a version of Bell's inequality for three qubits. In classical systems, the outcomes for pairs of coins are bound by the statistical constraint $C \geq 1$. Yet, quantum entanglement allows for the violation of this classical limit. Show that:

- (d) If Alice and Bob use entangled qubit pairs instead of classical coins, and if Alice's measurements are set at angles $0^\circ, 120^\circ$, and -120° (corresponding to A_1, A_2 , and A_3), while Bob's are at $180^\circ, -60^\circ$, and 60° (corresponding to B_1, B_2 , and B_3), they can achieve $C = 0.75$, violating the classical bound of $C \geq 1$.

10. Key Applications of Entanglement

Contents

10.1	Review of Preliminaries	246
10.1.1	Bell State Generation and Measurement	246
10.1.2	Conversion of Single-Qubit States	247
10.1.3	Conversion of Bell States	247
10.2	Superdense Coding	248
10.2.1	Introduction	248
10.2.2	Step-by-Step Analysis	248
10.2.3	* Further Exploration	250
10.3	Quantum Teleportation	251
10.3.1	Introduction	251
10.3.2	Basis State Approach	252
10.3.3	* Logic Operation Approach	254
10.3.4	* Bell Measurement Approach	256
10.3.5	* Further Exploration	259
10.4	Entanglement Swapping	260
10.4.1	Introduction	260
10.4.2	* Step-by-Step Analysis	261
10.4.3	* Further Exploration	264
10.5	Quantum Gate Teleportation	264
10.5.1	* Single-Qubit Gate Teleportation	265
10.5.2	* Teleportation of Two-Qubit Gates	271
10.5.3	* Further Exploration	276
10.6	E91 Quantum Key Distribution Protocol	277
10.6.1	Procedure of the E91 QKD Protocol	278
10.6.2	Comparison to the BB84 Protocol	280
10.6.3	* Further Exploration	281
10.7	Summary and Conclusions	282
	Problem Set 10	283

Quantum entanglement underpins a myriad of applications in quantum computing and quantum communication. These applications range from efficient quantum teleportation and superdense coding, which facilitate communication, to resilient quantum information processing using error-correcting block codes. As research in these areas advances, we can expect a surge of even more innovative applications that leverage the remarkable capabilities of quantum entanglement.

Application	Purpose	Practical Uses
Superdense coding	Allows the transmission of two classical bits using a single entangled quantum bit.	Efficient quantum communication.
Quantum Teleportation	Allows the transfer of quantum information from one location to another, with the help of two classical bits and entanglement.	Quantum communication. Foundation for other entanglement-based applications.
Entanglement swapping	Allows entanglement to be shared between systems that have never interacted.	Creation of quantum networks and long-distance quantum communication.
Quantum gate teleportation	Facilitates the transfer of quantum gates or operations rather than states.	Extends the range of quantum operations. Applications in fault-tolerant computing.
E91 QKD	A quantum key distribution protocol that exploits the principles of quantum entanglement to guarantee secure communication.	Cryptographically secure communication in quantum networks.

Table 10.1: Summary of Key Applications of Quantum Entanglement

In this long chapter, we will explore a number of fundamental applications of quantum entanglement as listed in Table 10.1. Among these, quantum teleportation is the most fundamental, and we will delve into it in extensive detail. However, we will begin our journey with superdense coding, as it is the most straightforward to understand given its operation on the basis states of only two qubits.

10.1 Review of Preliminaries

This section provides a review of some preliminaries necessary for the exploration of quantum entanglement-based applications.

10.1.1 Bell State Generation and Measurement

Bell states represent maximally entangled two-qubit systems and provide a complete, orthogonal basis for the space of two-qubit states. For a detailed discussion of Bell

states, refer to Chapter 8.

For any $i, j \in \{0, 1\}$, these states are defined as follows:

$$|\beta_{00}\rangle \equiv |\Phi^+\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle), \quad (10.1a)$$

$$|\beta_{01}\rangle \equiv |\Psi^+\rangle = \frac{1}{\sqrt{2}} (|01\rangle + |10\rangle), \quad (10.1b)$$

$$|\beta_{10}\rangle \equiv |\Phi^-\rangle = \frac{1}{\sqrt{2}} (|00\rangle - |11\rangle), \quad (10.1c)$$

$$|\beta_{11}\rangle \equiv |\Psi^-\rangle = \frac{1}{\sqrt{2}} (|01\rangle - |10\rangle). \quad (10.1d)$$

The generation and analysis of Bell states form the basis of many quantum algorithms and protocols. The following figure illustrates the quantum circuits for a Bell state generator and a Bell state analyzer (also referred to as Bell measurement), respectively.

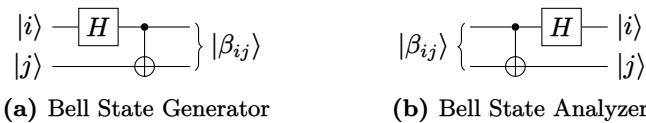


Figure 10.1: Bell State Generator and Analyzer

10.1.2 Conversion of Single-Qubit States

Table 10.2 shows how “correction gates” convert a qubit state $|\psi_{ij}\rangle$ into the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. These conversions are reversible with the same gates, enabling state interchangeability, ignoring a global phase factor.

Index ij	Qubit State $ \psi_{ij}\rangle$	Correction Gates
00	$ \psi_{00}\rangle = \alpha 0\rangle + \beta 1\rangle$	I (identity, or none)
01	$ \psi_{01}\rangle = \alpha 1\rangle + \beta 0\rangle$	X (bit-flip)
10	$ \psi_{10}\rangle = \alpha 0\rangle - \beta 1\rangle$	Z (phase-flip)
11	$ \psi_{11}\rangle = \alpha 1\rangle - \beta 0\rangle$	$X \& Z$ or iY (bit and phase flip)

Table 10.2: Correction Gates for General Qubit States

These relations can be represented by the following equations:

$$|\psi_{00}\rangle \equiv |\psi\rangle, \quad (10.2a)$$

$$|\psi_{01}\rangle = X|\psi\rangle, \quad |\psi\rangle = X|\psi_{01}\rangle, \quad (10.2b)$$

$$|\psi_{10}\rangle = Z|\psi\rangle, \quad |\psi\rangle = Z|\psi_{10}\rangle, \quad (10.2c)$$

$$|\psi_{11}\rangle = ZX|\psi\rangle, \quad |\psi\rangle = ZX|\psi_{11}\rangle. \quad (10.2d)$$

10.1.3 Conversion of Bell States

Similarly, Pauli gates applied to *either one* of the qubits in a Bell pair can transform the state $|\beta_{ij}\rangle$ into $|\beta_{00}\rangle$ or vice versa, up to a global phase factor. The transformation

depends on the index ij , as summarized in Table 10.3. For a full treatment of this topic, see § 8.5.

Index ij	Bell State	Conversion Gates
00	$ \beta_{00}\rangle \equiv \Phi^+\rangle = \frac{1}{\sqrt{2}}(00\rangle + 11\rangle)$	I (identity, or none)
01	$ \beta_{01}\rangle \equiv \Psi^+\rangle = \frac{1}{\sqrt{2}}(01\rangle + 10\rangle)$	X (bit-flip)
10	$ \beta_{10}\rangle \equiv \Phi^-\rangle = \frac{1}{\sqrt{2}}(00\rangle - 11\rangle)$	Z (phase-flip)
11	$ \beta_{11}\rangle \equiv \Psi^-\rangle = \frac{1}{\sqrt{2}}(01\rangle - 10\rangle)$	X & Z or iY (bit and phase flip)

Table 10.3: Conversion Gates for Bell States

These relations can be represented by the following equations:

$$|\beta_{01}\rangle = (I \otimes X) |\beta_{00}\rangle, \quad |\beta_{00}\rangle = (I \otimes X) |\beta_{01}\rangle, \quad (10.3a)$$

$$|\beta_{01}\rangle = (X \otimes I) |\beta_{00}\rangle, \quad |\beta_{00}\rangle = (X \otimes I) |\beta_{01}\rangle, \quad (10.3b)$$

$$|\beta_{10}\rangle = (I \otimes Z) |\beta_{00}\rangle, \quad |\beta_{00}\rangle = (I \otimes Z) |\beta_{10}\rangle, \quad (10.3c)$$

$$|\beta_{10}\rangle = (Z \otimes I) |\beta_{00}\rangle, \quad |\beta_{00}\rangle = (Z \otimes I) |\beta_{10}\rangle, \quad (10.3d)$$

$$|\beta_{11}\rangle = (I \otimes ZX) |\beta_{00}\rangle, \quad |\beta_{00}\rangle = (I \otimes ZX) |\beta_{11}\rangle, \quad (10.3e)$$

$$|\beta_{11}\rangle = (ZX \otimes I) |\beta_{00}\rangle, \quad |\beta_{00}\rangle = (XZ \otimes I) |\beta_{11}\rangle. \quad (10.3f)$$

10.2 Superdense Coding

10.2.1 Introduction

Basic information theory dictates that by sending n bits, one cannot communicate more than n bits of information. Superdense coding, however, is a quantum communication protocol that defies this limit by enabling the transmission of two classical bits of information using only a single qubit. This remarkable achievement is made possible through the use of quantum entanglement, local operations, and classical communication channels. In the absence of prior entanglement, the sender is limited to conveying no more than one bit of information per qubit, a constraint fundamentally outlined in Holevo's Theorem (§ 13.3.2). Superdense coding opens the door to potential advancements in communication efficiency, secure data transmission, and distributed quantum computing.

Superdense Coding: Core Intuition

Superdense coding utilizes: (1) the transformation of any Bell pair into another by applying a Pauli gate to one of the qubits locally (see § 10.1.3); and (2) the orthogonality of Bell states, which enables their perfect discrimination through measurement, as discussed in § 8.2.

10.2.2 Step-by-Step Analysis

1 Bell State Preparation

The procedure of superdense coding encompasses four primary steps, as depicted in

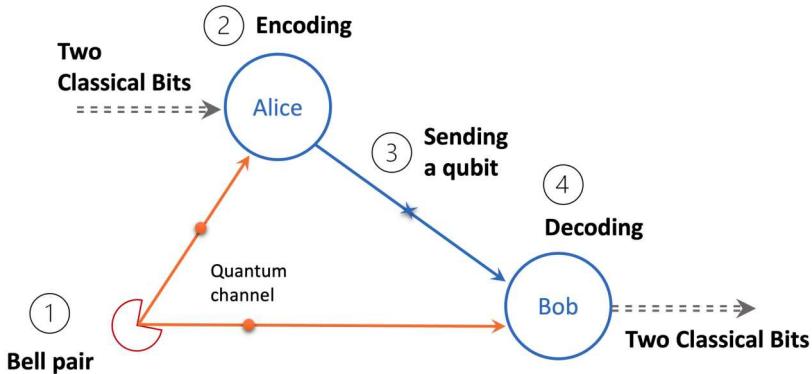


Figure 10.2: Basic Process of Superdense Coding

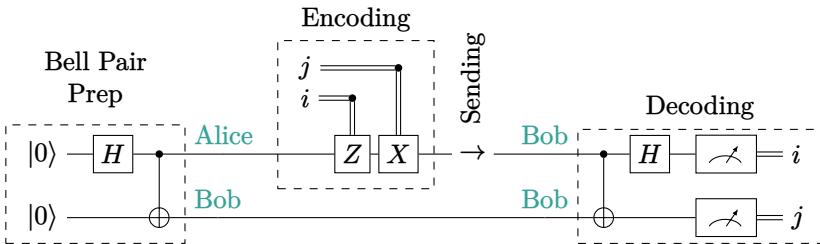


Figure 10.3: Basic Quantum Circuit for Superdense Coding

Fig. 10.2. The corresponding quantum circuit is shown in Fig. 10.3. Alice and Bob share an entangled pair of qubits, pre-established by Alice, Bob, or a third party. The initial state vector of the system is the Bell state:

$$|\Psi_0\rangle = |\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \quad (10.4)$$

2 Encoding

Alice wants to send two classical bits of information, ij , to Bob. Depending on the two bits she wants to send, Alice applies some Pauli gates to her qubit, according to Table 10.3. For '00', no gates are applied. For '01', the Z gate is applied. For '10', the X gate is applied. For '11', both the X and Z gates are applied. These gates transform the original Bell state $|\Phi^+\rangle \equiv |\beta_{00}\rangle$ to another Bell state, which is the system state after encoding:

Bits to Send (ij)	Encoding Gates	Encoded State (New Bell State)
00	I (identity)	$ \Phi^+\rangle \equiv \beta_{00}\rangle = \frac{1}{\sqrt{2}}(00\rangle + 11\rangle)$
01	X (bit-flip)	$ \Psi^+\rangle \equiv \beta_{01}\rangle = \frac{1}{\sqrt{2}}(10\rangle + 01\rangle)$
10	Z (phase-flip)	$ \Phi^-\rangle \equiv \beta_{10}\rangle = \frac{1}{\sqrt{2}}(00\rangle - 11\rangle)$
11	X & Z or iY	$ \Psi^-\rangle \equiv \beta_{11}\rangle = \frac{1}{\sqrt{2}}(01\rangle - 10\rangle)$

3 Decoding

Alice sends her qubit to Bob. Bob now has both qubits. He applies a CNOT gate with the received qubit as control and his own as the target, followed by a Hadamard gate on the control qubit. Bob then measures both qubits in the computational basis. Bob thus effectively performs a Bell measurement, and recovers the two classical bits Alice intended to send.

Alice's Bits	Encoded Bell State	Bob's Bits
00	$ \Phi^+\rangle \equiv \beta_{00}\rangle = \frac{1}{\sqrt{2}}(00\rangle + 11\rangle)$	00
01	$ \Psi^+\rangle \equiv \beta_{01}\rangle = \frac{1}{\sqrt{2}}(10\rangle + 01\rangle)$	01
10	$ \Phi^-\rangle \equiv \beta_{10}\rangle = \frac{1}{\sqrt{2}}(00\rangle - 11\rangle)$	10
11	$ \Psi^-\rangle \equiv \beta_{11}\rangle = \frac{1}{\sqrt{2}}(01\rangle - 10\rangle)$	11

Exercise 10.1 Explain how to perform superdense coding if Alice and Bob share a different Bell state at the start (e.g., $|\Phi^-\rangle$, $|\Psi^+\rangle$, or $|\Psi^-\rangle$). What operations should Alice use to encode her two bits of information?

Why is superdense coding potentially more efficient than classical transmission if both methods involve two bits? In superdense coding, one of the qubits can be shared with the receiver before the message is decided. So, when it's time to communicate the message, only one qubit needs to be sent. This method converts high-latency bandwidth into low-latency bandwidth. The receiver measures two qubits, but the first one could have been received long before the message was decided. This technique can effectively double the communication rate, given enough pre-shared entanglement.

10.2.3 *Further Exploration

There are variations and extensions to the concept of superdense coding. These variations encompass the type and dimension of the quantum states utilized, the number of participants, and the operations applied to the states. Aspired learners are encouraged to delve into these topics to gain a comprehensive understanding.

1. Higher-dimensional Systems and Quantum Networks [50]: The protocol of superdense coding has been extended beyond mere qubits. Its evolution is discussed from discrete and continuous variables to quantum networks. This progress report elucidates the foundational theoretical principles behind quantum dense coding and spans its variant protocols to its applications in quantum secure communication.
2. Probabilistic Superdense Coding [71]: The feasibility of performing superdense coding with nonmaximally entangled states is examined in this study. The authors highlight that it's possible to transmit two classical bits probabilistically by sending just one qubit.
3. Distributed Quantum Dense Coding [32]: In this paper, the notion of distributed quantum dense coding is introduced. This is a generalization of quantum dense coding that incorporates more than one sender and receiver. The research proposes a classification scheme of quantum states based on their

utility in dense coding. The study particularly emphasizes that in the bipartite scenario, bound entanglement is not conducive for this endeavor.

10.3 Quantum Teleportation

Quantum teleportation facilitates the transfer of quantum information (the state of a qubit) from one point to another without physically transmitting the qubit itself. This remarkable accomplishment is realized via entanglement, classical communication channels, and local operations. Quantum teleportation has potential applications in secure communication and distributed quantum computing.

Quantum teleportation is the only known method capable of transferring quantum information between systems without data loss. Due to quantum superposition, it's impossible to measure and replicate all properties of a quantum system for reconstruction elsewhere. In quantum mechanics, a system exists in a combination of multiple potential states. Upon measurement, the system collapses to a single state, erasing information about other possibilities. Quantum teleportation, however, enables the replication of a qubit state, including its superposition, in another qubit. This process results in the destruction of the original state in the source qubit, thus adhering to the no-cloning theorem.

Quantum teleportation not only forms a key foundation in quantum computing, but also serves as an exemplary case study for contrasting and comprehending different analytical methods. If you aim to pioneer in the field of quantum computation or potentially conceive innovative algorithms similar to quantum teleportation, it is essential to have a comprehensive understanding of these varied methods. Through this exploration, our goal is to arm you with the analytical tools necessary for further pioneering and advancement in this domain.

10.3.1 Introduction

Quantum teleportation is a complex process that facilitates the transfer of quantum information from one point to another, without physically moving the quantum system itself. At its heart, this procedure encompasses four primary steps, as depicted in Fig. 10.4.

1. Establishing Entanglement: The sender (Alice) and the receiver (Bob) share an entangled quantum state, typically a pair of qubits in a Bell state.
2. Bell Measurement: Alice conducts a Bell measurement on her part of the entangled pair and the qubit she wishes to teleport, collapsing these qubits into one of four Bell states. This action annihilates the original state of the qubit Alice intends to teleport.
3. Classical Communication: Alice conveys the outcomes of her measurement to Bob via a classical communication channel.
4. State Reconstruction: On receiving this information, Bob applies a specific quantum gate to his part of the entangled pair, contingent on Alice's measurement result. This step reproduces the original quantum state of the qubit that Alice intended to teleport on Bob's qubit.

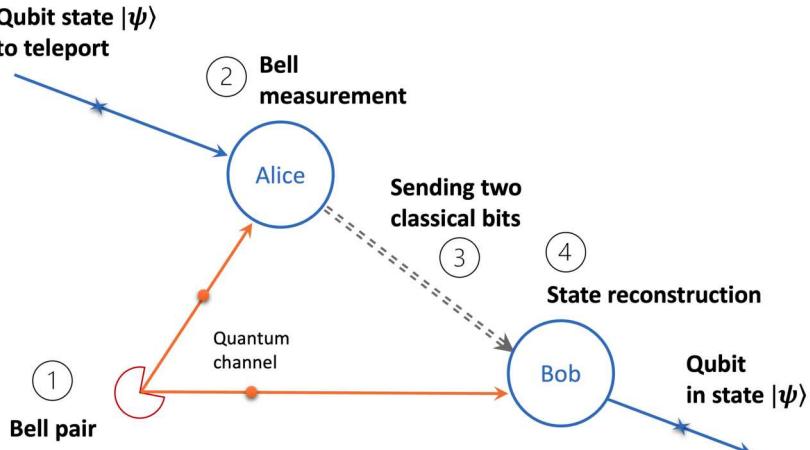


Figure 10.4: Basic Process of Quantum Teleportation

The fundamental procedure for quantum teleportation can be enacted by the quantum circuit shown in Fig. 10.5, which we will dissect in the forthcoming discussion.

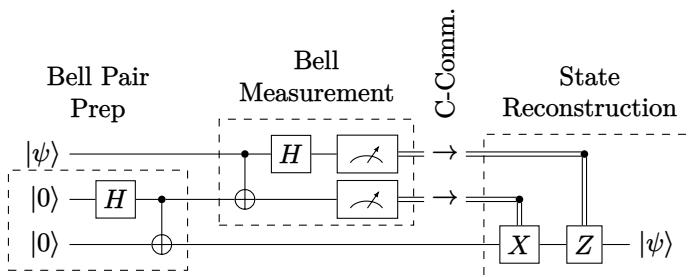


Figure 10.5: Fundamental Quantum Circuit for Teleportation

Several strategies exist for analyzing and comprehending quantum teleportation, and our objective is to explore three separate but interrelated approaches: the basis state approach, the logic operation approach, and the Bell measurement approach. Each of these methods showcases essential toolsets and provides unique insights into this foundational concept, thereby proving invaluable in the analysis of more advanced quantum algorithms.

10.3.2 Basis State Approach

This approach involves examining the teleportation circuit separately for each of the basis states $|0\rangle$ and $|1\rangle$. Utilizing the principle of linearity inherent to quantum transformations, we can subsequently combine these separate analyses to understand the teleportation of a general quantum state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, which Alice wishes to transmit.

This method allows us to derive the state of all three qubits at each step without invoking advanced concepts such as Bell measurements. However, a downside of this

approach is that the fundamental physical principles become somewhat obscured within routine algebraic manipulations.

1 Deferred Measurement Principle

To simplify the analysis, we will employ the principle of deferred measurement (introduced in § 7.4.5). This principle states that measurements can be shifted from an intermediate stage of a quantum circuit to the end. Furthermore, if the measurement results are used at any point in the circuit, the classically controlled operations can be replaced by quantum controlled operations.

Leveraging this principle, we can substitute classically controlled X and Z gates with quantum CX (i.e., CNOT) and CZ gates, as illustrated in Fig. 10.6. The subsequent analysis uses this version of the quantum teleportation circuit.

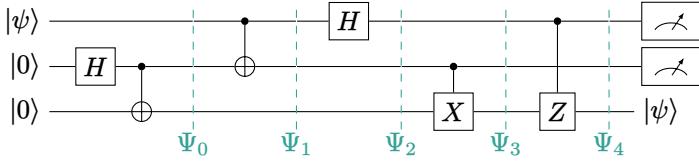


Figure 10.6: Quantum Circuit for Teleportation with Deferred Measurement

2 Establishing Entanglement

The initial state vector of the system after the Bell state preparation is:

$$|\Psi_0\rangle = \begin{cases} \frac{1}{\sqrt{2}}(|000\rangle + |011\rangle) & \text{for basis state } |0\rangle, \\ \frac{1}{\sqrt{2}}(|100\rangle + |111\rangle) & \text{for basis state } |1\rangle. \end{cases} \quad (10.5)$$

3 Alice Applies CNOT Gate

Now Alice applies a CNOT gate with her qubit (first qubit) as the control and her entangled qubit (second qubit) as the target:

$$|\Psi_1\rangle = \begin{cases} \frac{1}{\sqrt{2}}(|000\rangle + |011\rangle) & \text{for basis state } |0\rangle, \\ \frac{1}{\sqrt{2}}(|110\rangle + |101\rangle) & \text{for basis state } |1\rangle. \end{cases} \quad (10.6)$$

4 Alice Applies H Gate

Then Alice applies a Hadamard gate to her qubit:

$$|\Psi_2\rangle = \begin{cases} \frac{1}{2}(|000\rangle + |011\rangle + |100\rangle + |111\rangle) & \text{for basis state } |0\rangle, \\ \frac{1}{2}(|010\rangle + |001\rangle - |110\rangle - |101\rangle) & \text{for basis state } |1\rangle. \end{cases} \quad (10.7)$$

5 Bob Applies CX Gate

Then Alice measures the second qubit and Bob applies an X gate to the third qubit if the measurement result is 1. This is equivalent to applying a controlled- X (same as CNOT) gate, which flips the third qubit if the second qubit is 1.

$$|\Psi_3\rangle = \begin{cases} \frac{1}{2}(|000\rangle + |010\rangle + |100\rangle + |110\rangle) & \text{for basis state } |0\rangle, \\ \frac{1}{2}(|011\rangle + |001\rangle - |111\rangle - |101\rangle) & \text{for basis state } |1\rangle. \end{cases} \quad (10.8)$$

6 Bob Applies CZ Gate

Finally, Alice measures the first qubit and Bob applies a Z gate to the third qubit if the measurement result is 1. This is equivalent to applying a controlled- Z gate, which changes the sign of the state vector if both the second and third qubits are 1.

$$|\Psi_4\rangle = \begin{cases} \frac{1}{2}(|000\rangle + |010\rangle + |100\rangle + |110\rangle) & \text{for basis state } |0\rangle, \\ \frac{1}{2}(|011\rangle + |001\rangle + |111\rangle + |101\rangle) & \text{for basis state } |1\rangle, \end{cases} \quad (10.9)$$

$$= \begin{cases} \frac{1}{2}(|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \otimes |0\rangle & \text{for basis state } |0\rangle, \\ \frac{1}{2}(|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \otimes |1\rangle & \text{for basis state } |1\rangle. \end{cases} \quad (10.10)$$

If we use $|x\rangle$ to represent the basis states, where $x \in \{0, 1\}$, we can combine the above equation as:

$$|\Psi_4\rangle = \frac{1}{2}(|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \otimes |x\rangle. \quad (10.11)$$

At this stage, it becomes evident that the third qubit has become disentangled from the first two qubits, since its state is now a component of a product state. Specifically, it is in the state $|x\rangle$. Since $|x\rangle$ can represent any basis state and quantum transformations are linear by nature, Eq. 10.11 will still hold if we replace $|x\rangle$ with a general qubit state $|\psi\rangle$. In fact, $|\Psi_4\rangle = |+\rangle|+\rangle|\psi\rangle$. This indicates that Alice has successfully teleported her qubit state $|\psi\rangle$ to Bob.

10.3.3 *Logic Operation Approach

This method builds on the basis state approach but deviates by directly deriving equations for the combined state $|x\rangle$, rather than considering individual basis states. This is achieved by employing mathematical operations using the Boolean representation technique, similar to Boolean algebra used for classical logic gates.

The Boolean representation of quantum gates and circuits is detailed in § 7.3. Below is a brief summary.

With $|x\rangle$ and $|y\rangle$ representing $|0\rangle$ and $|1\rangle$, i.e., $x, y \in \{0, 1\}$, and $\bar{x} \equiv 1 - x$, $\bar{y} \equiv 1 - y$, the Boolean representation of some popular gates is summarized in Table 10.4.

Gate	Boolean Representation
X or NOT	$ x\rangle \rightarrow \bar{x}\rangle$
CNOT or CX	$ x, y\rangle \rightarrow x, x \oplus y\rangle$
Z	$ x\rangle \rightarrow (-1)^x x\rangle$
CZ	$ x, y\rangle \rightarrow (-1)^{x \cdot y} x, y\rangle$
H	$ x\rangle \rightarrow \frac{1}{\sqrt{2}} (0\rangle + (-1)^x 1\rangle)$
Y	$ x\rangle \rightarrow i(-1)^x \bar{x}\rangle$

Table 10.4: Boolean Representation of Selected Qubit Gates

Here, \oplus denotes the exclusive OR (or XOR) Boolean operator, which corresponds to bitwise modulo 2 addition. Below are some useful identities related to XOR:

$$x \oplus y = x + y \pmod{2} \quad (10.12a)$$

$$x \oplus 0 = 0 \oplus x = x \quad (10.12b)$$

$$x \oplus 1 = 1 \oplus x = \bar{x} \quad (10.12c)$$

$$\bar{x} \oplus 1 = 1 \oplus \bar{x} = x \quad (10.12d)$$

$$x \oplus x = 0 \quad (10.12e)$$

$$x \oplus \bar{x} = \bar{x} \oplus x = 1 \quad (10.12f)$$

$$\overline{x \oplus y} = \bar{x} \oplus y = x \oplus \bar{y} = 1 \oplus x \oplus y \quad (10.12g)$$

$$(-1)^{x \oplus y} = (-1)^{\bar{x} \oplus \bar{y}} = (-1)^{x+y} \quad (10.12h)$$

$$(-1)^{x \oplus \bar{y}} = (-1)^{\bar{x} \oplus y} = (-1)^{1+x+y}. \quad (10.12i)$$

With the above mathematical tool, let's revisit the steps from the basis states approach, referring to Fig. 10.6.

1 Establishing Entanglement

The initial state vector of the system is:

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}}(|x00\rangle + |x11\rangle). \quad (10.13)$$

2 Alice Applies CNOT Gate

Now Alice applies a CNOT gate with her qubit (first qubit) as the control and her entangled qubit (second qubit) as the target:

$$|\Psi_1\rangle = \frac{1}{\sqrt{2}}(|x(0 \oplus x)0\rangle + |x(1 \oplus x)1\rangle) \quad (10.14a)$$

$$= \frac{1}{\sqrt{2}}(|xx0\rangle + |x\bar{x}1\rangle). \quad (10.14b)$$



Did the CNOT operation entangle the first and second qubit? Yes. Even though $|\Psi_1\rangle = |x\rangle \otimes \frac{1}{\sqrt{2}}(|x0\rangle + |\bar{x}1\rangle)$ appears to be a product state, it is not. It represents $\frac{1}{\sqrt{2}}(\alpha|0\rangle(|00\rangle + |11\rangle) + \beta|1\rangle(|10\rangle + |01\rangle))$ which cannot be decomposed as the tensor product of $\alpha|0\rangle + \beta|1\rangle$ with another state.

3 Alice Applies H Gate

Then Alice applies a Hadamard gate to her qubit (first qubit). To make the equations shorter, let's rewrite $|\Psi_1\rangle$ as:

$$|\Psi_1\rangle = |x\rangle \otimes \frac{1}{\sqrt{2}}(|x0\rangle + |\bar{x}1\rangle). \quad (10.15)$$

After the application of the H gate:

$$|\Psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (-1)^x|1\rangle) \otimes \frac{1}{\sqrt{2}}(|x0\rangle + |\bar{x}1\rangle) \quad (10.16a)$$

$$= \frac{1}{2}(|0x0\rangle + |0\bar{x}1\rangle + (-1)^x|1x0\rangle + (-1)^x|1\bar{x}1\rangle). \quad (10.16b)$$

i

How many qubits are currently entangled? Three. Although Eq. 10.16a seems to represent a product state, the inclusion of the $(-1)^x$ factor renders it not a genuine product state.

4 Bob Applies CX Gate

Then Alice measures the second qubit and Bob applies an X gate to the third qubit if the measurement result is 1. Due to the deferred measurement theorem, this is equivalent to applying a CX (i.e., CNOT) gate.

$$|\Psi_3\rangle = \frac{1}{2}(|0x(x \oplus 0)\rangle + |0\bar{x}(\bar{x} \oplus 1)\rangle + (-1)^x |1x(x \oplus 0)\rangle + (-1)^x |1\bar{x}(\bar{x} \oplus 1)\rangle) \quad (10.17a)$$

$$= \frac{1}{2}(|0xx\rangle + |0\bar{x}\bar{x}\rangle + (-1)^x |1xx\rangle + (-1)^x |1\bar{x}\bar{x}\rangle). \quad (10.17b)$$

5 Bob Applies CZ Gate

Finally, Alice measures the first qubit and Bob applies a Z gate to the third qubit if the measurement result is 1:

$$|\Psi_4\rangle = \frac{1}{2}(|0xx\rangle + |0\bar{x}\bar{x}\rangle + (-1)^x(-1)^x |1xx\rangle + (-1)^x(-1)^x |1\bar{x}\bar{x}\rangle) \quad (10.18a)$$

$$= \frac{1}{2}(|0xx\rangle + |0\bar{x}\bar{x}\rangle + |1xx\rangle + |1\bar{x}\bar{x}\rangle) \quad (10.18b)$$

$$= \frac{1}{2}(|0\rangle + |1\rangle) \otimes (|x\rangle + |\bar{x}\rangle) \otimes |x\rangle \quad (10.18c)$$

$$= \frac{1}{2}(|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \otimes |x\rangle. \quad (10.18d)$$

At this stage, it becomes evident that the three qubits become disentangled, and the third qubit is $|x\rangle$. Alice has successfully teleported her qubit state $|\psi\rangle$ to Bob.

Exercise 10.2 Try to replicate the above analysis steps. This time, however, assume that Alice and Bob share a different Bell state, $|\Psi^-\rangle$, at the start.

10.3.4 *Bell Measurement Approach

We will now examine the quantum teleportation process by utilizing Bell measurement, discussed § 8.4, as an analytical tool. Despite incorporating the advanced concept of Bell measurement, this approach provides the clearest physical intuition about the construction of the teleportation circuit in Fig. 10.5.

1 Establishing Entanglement

Alice wants to send a quantum state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ to Bob. They share an entangled pair of qubits in the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. The initial state vector of the system can be written as:

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle) \otimes (|00\rangle + |11\rangle) \quad (10.19a)$$

$$= \frac{1}{\sqrt{2}}(\alpha|000\rangle + \alpha|011\rangle + \beta|100\rangle + \beta|111\rangle) \quad (10.19b)$$

$$= \frac{1}{\sqrt{2}}(\alpha|00\rangle + \beta|10\rangle)|0\rangle + \frac{1}{\sqrt{2}}(\alpha|01\rangle + \beta|11\rangle)|1\rangle. \quad (10.19c)$$

In the last step, we have singled out the 3rd qubit since we are going to measure the first two qubits in the next step.

2 Bell Measurement

Alice applies a CNOT gate with her qubit as the control and the entangled qubit as the target. Then she applies a Hadamard gate to her qubit. Finally, she measures the two qubits, effectively performing a Bell measurement on them.

The four Bell states form an orthonormal basis for two-qubit states. By performing Bell measurements, Alice projects the two-qubit states onto the Bell states.

The measurement operators, for the overall system of three qubits, are given by:

$$M_{ij} = |\beta_{ij}\rangle\langle\beta_{ij}| \otimes I, \quad (10.20)$$

where $i, j \in \{0, 1\}$, β_{ij} are the Bell states given by Eq. 10.1, and I is the identity operator applied to the 3rd qubit.

After the measurement, depending on the measurement result (i, j) , the system will be in the state:

$$|\Psi_1\rangle = M_{ij}|\Psi_0\rangle \quad (10.21a)$$

$$= \frac{1}{\sqrt{2}}(|\beta_{ij}\rangle\langle\beta_{ij}| \otimes I)((\alpha|00\rangle + \beta|10\rangle)|0\rangle + (\alpha|01\rangle + \beta|11\rangle)|1\rangle) \quad (10.21b)$$

$$\begin{aligned} &= \frac{1}{\sqrt{2}}|\beta_{ij}\rangle\langle\beta_{ij}|(\alpha|00\rangle + \beta|10\rangle)|0\rangle \\ &\quad + \frac{1}{\sqrt{2}}|\beta_{ij}\rangle\langle\beta_{ij}|(\alpha|01\rangle + \beta|11\rangle)|1\rangle \end{aligned} \quad (10.21c)$$

$$= |\beta_{ij}\rangle \otimes |\psi_{ij}\rangle, \quad (10.21d)$$

where $|\psi_{ij}\rangle$ is defined in Table 10.2, copied below for convenience:

$$|\psi_{00}\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (10.22a)$$

$$|\psi_{01}\rangle = \alpha|1\rangle + \beta|0\rangle, \quad (10.22b)$$

$$|\psi_{10}\rangle = \alpha|0\rangle - \beta|1\rangle, \quad (10.22c)$$

$$|\psi_{11}\rangle = \alpha|1\rangle - \beta|0\rangle. \quad (10.22d)$$



How is $|\Psi_1\rangle$ in this equation related to the $|\Psi_1\rangle$ in Eq. 10.6? The latter is a linear combination of the four cases of the former - see Eq. 10.23.

3 Classical Communication

Alice transmits the measurement outcomes (i, j , representing two bits of information) to Bob through a classical communication channel. It is important to note that the original qubit state encapsulates more information than merely two bits; it's characterized by two complex numbers, α and β . Furthermore, exact values cannot be obtained from a single measurement, as performing a measurement invariably disrupts the original quantum state.

4 State Reconstruction

Based on the two bits (i, j) resulting from Alice's measurements, Bob will apply the necessary correction gates to his qubit: a Z gate is applied if $i = 1$, and an X gate is applied if $j = 1$.

Outcome ij	Intermediate Qubit State	Correction Gates
00	$\alpha 0\rangle + \beta 1\rangle$	I (None)
01	$\alpha 1\rangle + \beta 0\rangle$	X
10	$\alpha 0\rangle - \beta 1\rangle$	Z
11	$\alpha 1\rangle - \beta 0\rangle$	X & Z , or iY

The final state vector of Bob's qubit becomes $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$. Now Bob's qubit is in the same state as Alice's initial qubit, completing the quantum teleportation process.

i For measurement outcome '11', both an X and a Z gate are applied. Their sequence does not matter in this case, as swapping their order only incurs a global phase shift ($XZ = -ZX$), which is unobservable in quantum mechanics. However, if subsequent stages involve interference with other qubits, the order may become crucial due to the potential impact on relative phases. In such scenarios, the correct sequence is X before Z , i.e., in operator terms, ZX .

Exercise 10.3 Try to replicate the above analysis steps. This time, however, assume that Alice and Bob share a different Bell state, specifically $|\Psi^-\rangle$, at the outset. What differences do you observe from the initial scenario?

Quantum Teleportation: Core Summary

This analysis captures the essence of quantum teleportation. The initial state vector can be decomposed into the orthonormal Bell basis as follows:

$$\begin{aligned} |\Psi_0\rangle &= \frac{1}{\sqrt{2}}(\alpha |0\rangle + \beta |1\rangle) \otimes (|00\rangle + |11\rangle), \\ &= \frac{1}{2} \sum_{i,j \in \{0,1\}} |\beta_{ij}\rangle \otimes |\psi_{ij}\rangle. \end{aligned} \quad (10.23)$$

where $|\beta_{ij}\rangle$ represents the Bell states, as defined in Eq. 10.1, and $|\psi_{ij}\rangle$ is defined in Table 10.2 and Eq. 10.22a.

When Alice measures the first two qubits in the Bell basis $\{\beta_{ij}\}$, each of the Bell states occurs with equal probability. If she obtains $|\beta_{ij}\rangle$ (where $i, j \in \{0, 1\}$), the state of the system collapses to $|\beta_{ij}\rangle \otimes |\psi_{ij}\rangle$, and the third qubit takes on the state $|\psi_{ij}\rangle$.

The third qubit can then be adjusted based on Alice's measurement results (i, j) to reconstruct the original state $\alpha|0\rangle + \beta|1\rangle$.

10.3.5 *Further Exploration

The concept of quantum teleportation extends beyond the basic protocol discussed in this text. Many variations and enhancements center around the nature and dimensionality of the quantum states used, the number of parties involved, and the types of operations executed on the states. For those seeking to deepen their understanding, the following topics are suggested for further exploration:

1. Quantum Teleportation of Composite Systems and Entanglement [7]: : This topic deals with the teleportation of composite systems, which can involve multiple entangled qubits. Understanding this concept is fundamental in advancing complex quantum information processing protocols.
2. Superdense Teleportation [47]: : This variant combines principles of quantum teleportation and superdense coding to enable the transfer of more information than either standalone protocol. This involves the teleportation of a two-qubit state using two classical bits and a shared entangled pair of qubits.
3. Higher-dimensional Quantum Teleportation [26]: : While the fundamental quantum teleportation protocol involves qubits, researchers have broadened this framework to include qudits. Higher-dimensional quantum systems can potentially provide richer quantum information and more robust entanglement.
4. Multiparty Quantum Teleportation [0]: : Beyond the traditional two-party setup, quantum teleportation can be extended to multiple parties. This development could involve numerous participants each transmitting information through a shared multipartite entangled state.
5. Teleportation of an Arbitrary Two-Qubit State [75]: : This work expands the scope of quantum teleportation to two-qubit states, linking this to multipartite entanglement. The insights provided enhance the understanding of complex quantum information transfer and potential improvements in teleportation protocols.
6. Continuous-Variable Quantum Teleportation [44]: : This variation uses continuous-variable systems, such as the position and momentum of a particle. As these systems can assume any value within a certain range, this approach often involves quantum states of light and can potentially provide higher fidelity in teleportation.

10.4 Entanglement Swapping

Entanglement swapping, also known as quantum relay, is a fundamental procedure in quantum information science that allows two quantum systems to become entangled without direct interaction, purely through the use of entanglement and local operations. It plays an instrumental role in quantum repeaters and quantum networks, enhancing secure long-distance communication and creating a web of quantum devices.

Entanglement swapping is a powerful technique in overcoming the physical limitations inherent in creating direct entanglement between distant particles. Entanglement is typically established by allowing two quantum systems to interact directly. However, as the distance between the two systems grows, the feasibility of this interaction rapidly diminishes due to environmental interference. Entanglement swapping breaks down the barrier of distance, allowing quantum systems to become entangled without the necessity for direct interaction.

Understanding entanglement swapping is essential for those delving into the field of quantum networking and quantum cryptography. Its concept, combined with other quantum protocols such as quantum teleportation, enables the creation of secure and efficient quantum communication systems. In this section, we aim to shed light on the intricacies of entanglement swapping and its pivotal role in quantum information science.

10.4.1 Introduction

Entanglement swapping is a procedure that leads to the entanglement of two quantum systems that have never interacted before. It is executed through a sequence of steps as outlined in Fig. 10.7.

1. **Initial Entanglement:** Two pairs of entangled quantum states are created, typically qubits in a Bell state. The first pair is shared between Alice and a mediator (Charlie), and the second pair is shared between Charlie and Bob.
2. **Bell Measurement:** Charlie conducts a Bell measurement on his two qubits from the entangled pairs, collapsing these qubits into one of four Bell states.
3. **Classical Communication:** Charlie communicates the outcomes of his measurement to Bob via a classical communication channel.
4. **State Reconstruction:** Upon receiving Charlie's measurement result, Bob performs certain quantum operations on his qubit depending on the result. This step results in Alice's and Bob's qubits becoming entangled, thereby completing the entanglement swapping process.



In our example, the state reconstruction is done on Bob's qubit. Depending on the specific requirements of the quantum operation or communication, the state reconstruction can be performed on Alice's qubit, Bob's, or both. The choice of where to perform the state reconstruction can have implications for the efficiency, security, and other characteristics of these quantum protocols.

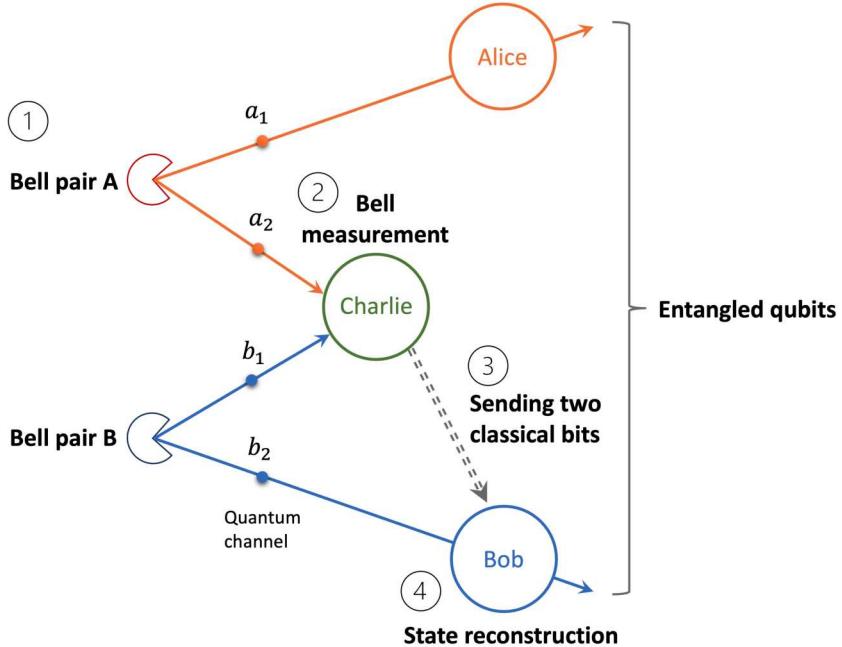


Figure 10.7: Basic Process of Entanglement Swapping

The fundamental procedure for entanglement swapping can be represented by the quantum circuit shown in Fig. 10.8, which we will examine in more detail in the following discussion.

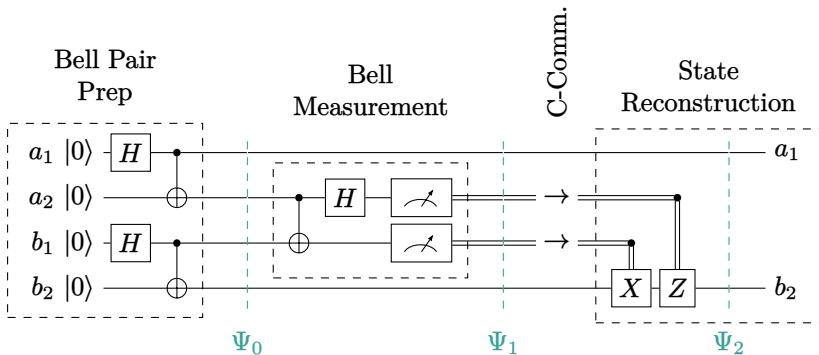


Figure 10.8: Fundamental Quantum Circuit for Entanglement Swapping

10.4.2 *Step-by-Step Analysis

Let's delve into the entanglement swapping protocol by going through each step. This process involves creating entanglement between two qubits without direct interaction, utilizing local operations, and classical communication.

1 Initial Entanglement

Consider two entangled pairs of qubits: Alice's pair A consisting of qubits a_1 and a_2 , and Bob's pair B consisting of qubits b_1 and b_2 . These pairs can be in any entangled state, but for simplicity, we'll consider them to be in the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.

The initial state vector of the system can be written as:

$$|\Psi_0\rangle_{a_1 a_2 b_1 b_2} = |\Phi^+\rangle_{a_1 a_2} \otimes |\Phi^+\rangle_{b_1 b_2} \quad (10.24a)$$

$$= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \quad (10.24b)$$

$$= \frac{1}{2}(|0000\rangle + |0011\rangle + |1100\rangle + |1111\rangle). \quad (10.24c)$$

For the purpose of the entanglement swapping protocol, we will treat a_2 and b_1 as one group, and a_1 and b_2 as another. This grouping aligns with the steps of the protocol, where a_2 and b_1 are sent to Charlie for a Bell measurement, while a_1 and b_2 become the final entangled pair. To reflect this in our notation, we will rearrange the order of the qubits in our state vector representation. Note that this is merely a mathematical adjustment and does not involve any physical changes to the qubits. In the new representation, the state vector is:

$$|\Psi_0\rangle_{a_2 b_1 a_1 b_2} = \frac{1}{2}(|0000\rangle + |0101\rangle + |1010\rangle + |1111\rangle). \quad (10.25)$$

2 Bell Measurement

Alice and Bob each send one qubit, a_2 and b_1 respectively, to Charlie. Upon receiving the qubits a_2 and b_1 , Charlie performs a Bell measurement. This measurement process is achieved by first applying a CNOT gate with a_2 as the control qubit and b_1 as the target, followed by a Hadamard gate on a_2 . This transforms the two-qubit state into a superposition of the Bell states. Subsequently, Charlie measures both qubits in the computational basis. For simplicity in this analysis, we treat Charlie's CNOT gate, Hadamard gate, and measurement devices as a single unit, which we refer to as the Bell-measurement unit.

The Bell measurement projects the two-qubit state onto one of the Bell states $|\beta_{ij}\rangle$. The measurement operators are:

$$M_{ij} = |\beta_{ij}\rangle\langle\beta_{ij}|_{a_2 b_1} \otimes I_{a_1 b_2}. \quad (10.26)$$

After the measurement, depending on the result, the system will be in one of the following states:

$$|\Psi_1\rangle_{ij} = M_{ij} |\Psi_0\rangle_{a_2 b_1 a_1 b_2}. \quad (10.27)$$

The result of this equation has the following form:

$$|\Psi_1\rangle_{ij} = |\beta_{ij}\rangle_{a_2 b_1} \otimes |\psi\rangle_{a_1 b_2}. \quad (10.28)$$

After some algebraic manipulation, we obtain $|\psi\rangle_{a_1 b_2}$ as in the following table:

Outcome ij	State of a_2 and b_1 ($ \beta_{ij}\rangle_{a_2 b_1}$)	State of a_1 and b_2 ($ \psi\rangle_{a_1 b_2}$)
00	$ \beta_{00}\rangle \equiv \Phi^+\rangle = \frac{1}{\sqrt{2}}(00\rangle + 11\rangle)$	$\frac{1}{\sqrt{2}}(00\rangle + 11\rangle)$
01	$ \beta_{01}\rangle \equiv \Psi^+\rangle = \frac{1}{\sqrt{2}}(01\rangle + 10\rangle)$	$\frac{1}{\sqrt{2}}(01\rangle + 10\rangle)$
10	$ \beta_{10}\rangle \equiv \Phi^-\rangle = \frac{1}{\sqrt{2}}(00\rangle - 11\rangle)$	$\frac{1}{\sqrt{2}}(00\rangle - 11\rangle)$
11	$ \beta_{11}\rangle \equiv \Psi^-\rangle = \frac{1}{\sqrt{2}}(01\rangle - 10\rangle)$	$\frac{1}{\sqrt{2}}(01\rangle - 10\rangle)$

Examining the above table, we discover that after the measurement with outcome ij , qubits a_1 and b_2 are also in the Bell state $|\beta_{ij}\rangle$! This relationship, which is the heart of entanglement swapping, is captured by the following equation:

$$|\Psi_1\rangle_{ij} = |\beta_{ij}\rangle_{a_2 b_1} \otimes |\beta_{ij}\rangle_{a_1 b_2}. \quad (10.29)$$

3 Classical Communication

Even though qubits a_1 and b_2 are in the Bell state $|\beta_{ij}\rangle$, the measurement outcome ij appears randomly, limiting the usefulness of the entanglement. Next, Charlie transmits the measurement outcomes ij to Bob through a classical communication channel for him to make corrections to the Bell state.

4 State Reconstruction

To transform the state of qubits a_1 and b_2 from $|\beta_{ij}\rangle$ to $|\Psi_2\rangle = |\beta_{00}\rangle$, Bob applies the necessary correction gates to his qubit b_2 according to the two bits ij from Charlie's measurements: an X gate when $j = 1$ and a Z gate when $i = 1$.

Outcome ij	State of a_1 and b_2	Correction Gates
00	$ \beta_{00}\rangle \equiv \Phi^+\rangle = \frac{1}{\sqrt{2}}(00\rangle + 11\rangle)$	I (None)
01	$ \beta_{01}\rangle \equiv \Psi^+\rangle = \frac{1}{\sqrt{2}}(01\rangle + 10\rangle)$	X
10	$ \beta_{10}\rangle \equiv \Phi^-\rangle = \frac{1}{\sqrt{2}}(00\rangle - 11\rangle)$	Z
11	$ \beta_{11}\rangle \equiv \Psi^-\rangle = \frac{1}{\sqrt{2}}(01\rangle - 10\rangle)$	X & Z

Entanglement Swapping: Core Summary

The re-grouped initial state can be decomposed into the orthonormal Bell basis as follows:

$$\begin{aligned} |\Psi_0\rangle_{a_2 b_1 a_1 b_2} &= \frac{1}{2}(|0000\rangle + |0101\rangle + |1010\rangle + |1111\rangle), \\ &= \frac{1}{2} \sum_{i,j \in \{0,1\}} |\beta_{ij}\rangle_{a_2 b_1} \otimes |\beta_{ij}\rangle_{a_1 b_2}. \end{aligned} \quad (10.30)$$

When Charlie measures qubits a_2 and b_1 in the Bell basis $\{\beta_{ij}\}$, each of the Bell states occurs with equal probability. If he obtains $|\beta_{ij}\rangle$ (where $i, j \in \{0, 1\}$), the state of the system collapses to $|\beta_{ij}\rangle \otimes |\beta_{ij}\rangle$, and qubits a_1 and b_2 also take

on the state $|\beta_{ij}\rangle$. They can then be adjusted based on the measurement results (i, j) to reconstruct the original entangled state $|\beta_{00}\rangle$.

10.4.3 *Further Exploration

The concept of entanglement swapping, similar to quantum teleportation, can be extended beyond the basic protocol outlined in this text. Numerous variations and enhancements focus on the nature and dimensionality of the quantum states used, the number of parties involved, and the types of operations executed on the states. For those interested in deepening their understanding, the following topics are suggested for further exploration:

1. **Quantum Repeaters with Entanglement Swapping [30]:** Explores the use of entanglement swapping in quantum communication infrastructure. Quantum repeaters help overcome distance limitations in quantum communication by using entanglement swapping to establish entanglement between distant parties.
2. **Multi-party Entanglement Swapping [94]:** Involves more than two parties, with multiple mediators performing Bell measurements and classical communications. This potentially enables large-scale quantum networks with distributed entanglement.
3. **Higher-dimensional Entanglement Swapping [25]:** Extends entanglement swapping from qubits to qudits, or higher-dimensional quantum systems. This extension can potentially enhance quantum information capacity and the robustness of entanglement.
4. **Entanglement Swapping with Mixed States [63]:** While most entanglement swapping schemes use pure entangled states, variations that incorporate mixed states exist. These protocols, though more complex, may be more practical as mixed states are commonly encountered in real quantum systems due to decoherence and noise.
5. **Entanglement Swapping in Various Systems [95]:** This research explores entanglement swapping in pure and noisy systems, analyzing the relationship between initial and final state entanglement using metrics like concurrence and negativity. The study reveals that the initial state's entanglement can influence the final state's average entanglement, with measurement bases and entanglement intensity playing critical roles.

10.5 Quantum Gate Teleportation

Quantum gate teleportation refers to the process of performing a quantum operation (a quantum gate) on one or more qubits that are not locally accessible, leveraging the properties of entanglement and quantum teleportation.

Quantum gate teleportation can be viewed as entanglement-assisted computation, in contrast to the concept of entanglement-assisted communication found in standard quantum teleportation. The primary aim is to perform a computational operation during the teleportation process itself.

Quantum gate teleportation is particularly beneficial in scenarios where direct manipulation of the qubit could lead to decoherence or when the qubits are spatially

separated. It can also enable the realization of distributed quantum computing, where quantum gates can be performed on qubits located in different quantum computers, thus enabling more extensive and powerful quantum networks.

10.5.1 * Single-Qubit Gate Teleportation

We will first consider single-qubit gate teleportation, where the goal is to apply a single-qubit unitary operation U to a quantum state $|\psi\rangle$ during teleportation.

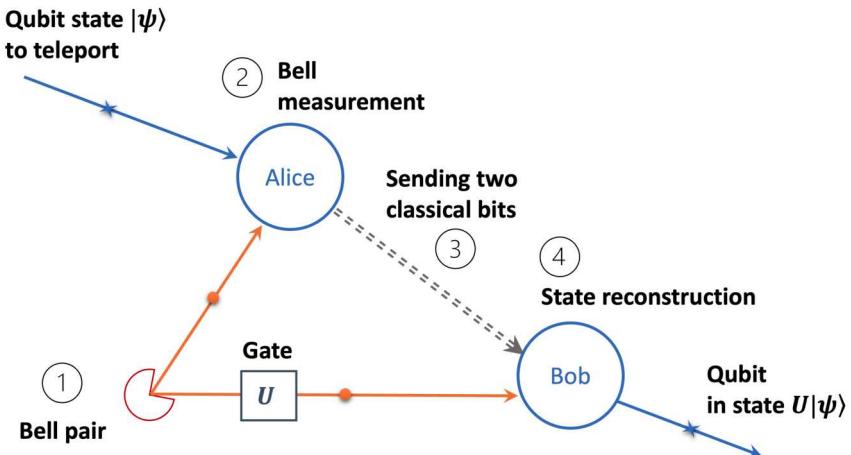


Figure 10.9: Basic Process of Quantum Gate Teleportation

Single-qubit gate teleportation follows a process similar to standard quantum teleportation, encompassing four main steps, as shown in Fig. 10.9. Instead of using the standard Bell state $|\Phi^+\rangle$, Alice and Bob begin with the shared state $(I \otimes U)|\Phi^+\rangle$, in which the unitary operation U is applied to Bob's part of the entangled pair. After Alice sends her measurement results, Bob performs correction operations, much like in standard teleportation, albeit with adjusted correction gates. The final state he obtains is $U|\psi\rangle$ instead of simply $|\psi\rangle$. This modification ensures that the teleportation procedure not only conveys the state $|\psi\rangle$ but also effectively applies the operation U to $|\psi\rangle$, resulting in the final state $U|\psi\rangle$ on Bob's end.

Figure 10.10 presents a quantum circuit for single-qubit gate teleportation that aligns with this procedure. We'll delve into this circuit in detail shortly.

1 Transition from Standard to Gate Teleportation

Let's now explore an intuitive connection between single-qubit gate teleportation and standard quantum teleportation. This exploration offers insight into the discovery process of protocols like gate teleportation. It also lays a foundation for understanding more complex protocols, such as CNOT gate teleportation, which we'll discuss later.

Refer to Fig. 10.11. The segment prior to the dashed visual separator is the standard teleportation circuit from Fig. 10.5. At this circuit's output, the third (bottom) qubit is in state $|\psi\rangle$. Upon application of a unitary gate U , the qubit's state becomes $U|\psi\rangle$. This final state mirrors that of single-qubit gate teleportation, wherein gate U is applied to the teleported $|\psi\rangle$ (as depicted in Fig. 10.10).

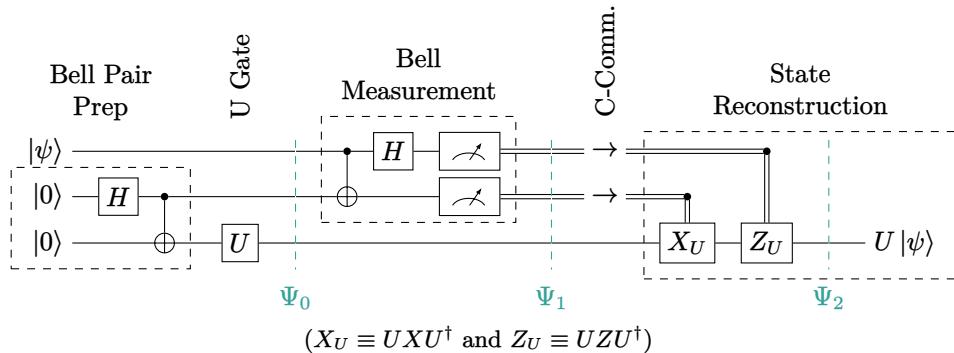


Figure 10.10: Quantum Circuit for Single-Qubit Gate Teleportation

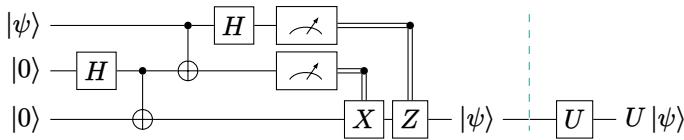


Figure 10.11: Gate U Applied Post-Teleportation

To transform the circuit in Fig. 10.11 into the gate teleportation circuit as shown in Fig. 10.10, we need to shift the gate U to the left of the controlled- Z and controlled- X gates. This transformation is facilitated by the concept of gate sequence equivalence, which is demonstrated in the following table:

Gate Sequence	Equivalent Seq.	Operator Equivalence
$[X] - [U]$	$-[U] - [X_U]$	$UX = UX(U^\dagger U) = (UXU^\dagger)U = X_U U$
$-[Z] - [U]$	$-[U] - [Z_U]$	$UZ = UZ(U^\dagger U) = (UZU^\dagger)U = Z_U U$

This concept of gate sequence equivalence clarifies the transformation of the correction gates from X and Z to $X_U \equiv UXU^\dagger$ and $Z_U \equiv UZU^\dagger$ respectively in Fig. 10.10.

i Although quantum circuit diagrams apply gates from left to right, the corresponding operator (matrix) multiplication operates in reverse order, from right to left.

Standard versus Gate Teleportation: Summary

In quantum teleportation, Alice and Bob share an entangled Bell state, $|\Phi^+\rangle$, allowing Alice to teleport an unknown state $|\psi\rangle$ to Bob. This is achieved by Alice measuring her qubits in the Bell basis, resulting in one of four possible outcomes which, on Bob's end, manifest as states $|\psi\rangle$, $X|\psi\rangle$, $Z|\psi\rangle$, or $ZX|\psi\rangle$. Bob corrects his state to $|\psi\rangle$ by applying the corresponding Pauli operator, I , X , Z , or XZ , informed by Alice's classical measurement.

Quantum gate teleportation extends this concept by teleporting a single-qubit gate U . Here, Alice and Bob share $(I \otimes U)|\Phi^+\rangle$ instead. The teleportation procedure leaves Bob with one of four states: $U|\psi\rangle$, $UX|\psi\rangle$, $UZ|\psi\rangle$, or $UZX|\psi\rangle$. Bob then corrects his state to $U|\psi\rangle$ by applying the corresponding operator, I , UXU^\dagger , UZU^\dagger , or $UXZU^\dagger$, depending on Alice's classical measurement.

Quantum teleportation is equivalent to quantum gate teleportation with $U = I$.

2 *Step-by-Step Analysis

While we've successfully derived the gate teleportation circuit by modifying the standard teleportation circuit, it's worthwhile to conduct a detailed, step-by-step examination of the gate teleportation process. This will serve to reinforce our understanding. Think of single-qubit gate teleportation as an extension of the Bell measurement approach to quantum teleportation, which we discussed in § 10.3.4. The equations that follow will closely mirror those found in the aforementioned section.

1. Initial State

Alice wants to send a quantum state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ to Bob, and have the state transformed by gate U during the process.

Alice and Bob share the state $(I \otimes U)|\Phi^+\rangle$, where U has been applied to Bob's part of the entangled pair. The initial state vector of the system can be written as:

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle) \otimes (I \otimes U)(|00\rangle + |11\rangle) \quad (10.31a)$$

$$= \frac{1}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle) \otimes (|0\rangle U|0\rangle + |1\rangle U|1\rangle) \quad (10.31b)$$

$$= \frac{1}{\sqrt{2}}(\alpha|00\rangle + \beta|10\rangle)U|0\rangle + \frac{1}{\sqrt{2}}(\alpha|01\rangle + \beta|11\rangle)U|1\rangle. \quad (10.31c)$$

Here in the last step, we have singled out the 3rd qubit since we are going to measure the first two qubits in the next step.

2. Bell Measurement

Alice applies a CNOT gate with her qubit as the control and the entangled qubit as the target. Then she applies a Hadamard gate to her qubit. Finally, she measures the two qubits, effectively performing a Bell measurement on them.

The four Bell states form an orthonormal basis for two-qubit states. By performing Bell measurements, Alice projects the two-qubit states onto the Bell states.

The measurement operators, for the overall system of three qubits, are given by:

$$M_{ij} = |\beta_{ij}\rangle\langle\beta_{ij}| \otimes I, \quad (10.32)$$

where $i, j \in \{0, 1\}$, and β_{ij} are the Bell states.

After the measurement, depending on the measurement results, the system will be at one of the following states:

$$|\Psi_1\rangle = M_{ij} |\Psi_0\rangle \quad (10.33a)$$

$$= \frac{1}{\sqrt{2}} (|\beta_{ij}\rangle\langle\beta_{ij}| \otimes I) ((\alpha|00\rangle + \beta|10\rangle)U|0\rangle + (\alpha|01\rangle + \beta|11\rangle)U|1\rangle) \quad (10.33b)$$

$$= \frac{1}{\sqrt{2}} |\beta_{ij}\rangle\langle\beta_{ij}| (\alpha|00\rangle + \beta|10\rangle)U|0\rangle + \frac{1}{\sqrt{2}} |\beta_{ij}\rangle\langle\beta_{ij}| (\alpha|01\rangle + \beta|11\rangle)U|1\rangle \quad (10.33c)$$

$$= |\beta_{ij}\rangle \otimes \begin{cases} \alpha U|0\rangle + \beta U|1\rangle & \text{for } i,j = 0,0 \\ \alpha U|1\rangle + \beta U|0\rangle & \text{for } i,j = 0,1 \\ \alpha U|0\rangle - \beta U|1\rangle & \text{for } i,j = 1,0 \\ \alpha U|1\rangle - \beta U|0\rangle & \text{for } i,j = 1,1 \end{cases} \quad (10.33d)$$

$$= |\beta_{ij}\rangle \otimes \begin{cases} U|\psi\rangle & \text{for } i,j = 0,0 \\ UX|\psi\rangle & \text{for } i,j = 0,1 \\ UZ|\psi\rangle & \text{for } i,j = 1,0 \\ UZX|\psi\rangle & \text{for } i,j = 1,1. \end{cases} \quad (10.33e)$$

3. Classical communication

Alice transmits the measurement outcomes ij , representing two bits of information, to Bob through a classical communication channel.

4. State Reconstruction

Based on the two bits ij resulting from Alice's measurements, Bob will apply the necessary correction gates to his qubit as shown in the following table.

ij	Qubit State	Correction Gates
00	$U \psi\rangle$	I or None
01	$UX \psi\rangle$	UXU^\dagger
10	$UZ \psi\rangle$	UZU^\dagger
11	$UZX \psi\rangle$	$UXZU^\dagger$

Consider the case where $ij = 01$. Bob's initial state is $UX|\psi\rangle$, and we aim for it to become $U|\psi\rangle$ after correction. This is achieved by applying UXU^\dagger , as $UXU^\dagger UX|\psi\rangle = UXU|\psi\rangle = U|\psi\rangle$, given $X^2 = I$.

After correction, Bob's qubit resides in state $U|\psi\rangle$. This illustrates gate teleportation, where the gate U effect is transferred to Bob's qubit without direct application to $|\psi\rangle$.

Exercise 10.4 In quantum gate teleportation, the U gate is typically applied before the Bell measurement step. However, since the U gate is applied to a different qubit from those involved in the Bell measurement, its application can be delayed until after the measurement. This modified sequence of operations is illustrated in Fig. 10.12. To confirm this modification is valid, repeat the analysis as done above for the regular sequence of operations.

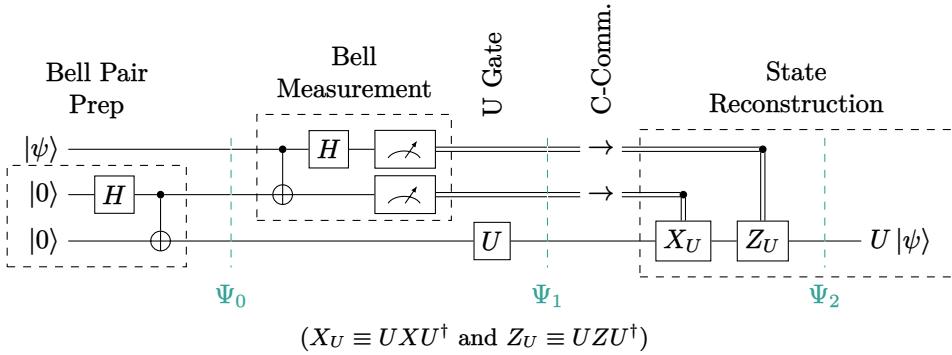


Figure 10.12: Single-Qubit Gate Teleportation with Deferred U Gate

Quantum Gate Teleportation: Core Summary

The initial state vector can be decomposed into the orthonormal Bell basis as follows:

$$\begin{aligned} |\Psi_0\rangle &= \frac{1}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle) \otimes (I \otimes U)(|00\rangle + |11\rangle), \\ &= \frac{1}{2} \sum_{i,j \in \{0,1\}} |\beta_{ij}\rangle \otimes U|\psi_{ij}\rangle, \end{aligned} \quad (10.34)$$

where $|\beta_{ij}\rangle$ represents the Bell states, and $|\psi_{ij}\rangle$ is defined in Table 10.2 and Eq. 10.22a.

When Alice measures the first two qubits in the Bell basis $\{\beta_{ij}\}$, each of the Bell states occurs with equal probability. If she obtains $|\beta_{ij}\rangle$ (where $i, j \in \{0, 1\}$), the state of the system collapses to $|\beta_{ij}\rangle \otimes U|\psi_{ij}\rangle$, and the third qubit takes on the state $U|\psi_{ij}\rangle$. It can then be adjusted based on Alice's measurement results (i, j) to reconstruct the state $U(\alpha|0\rangle + \beta|1\rangle)$.

3 * Teleportation of Single-Qubit Gates in the Clifford Group

You might wonder why we would teleport a gate U if the correction gates UXU^\dagger and UZU^\dagger also contain U . It might seem more straightforward to teleport $|\psi\rangle$ and then directly apply U . However, the process becomes simplified when U belongs to the Clifford group.

By definition, a single-qubit Clifford gate maps any Pauli operator to another Pauli operator under conjugation (up to a phase). This means that if U is in the Clifford group, UPU^\dagger is a Pauli operator for any Pauli operator P .

There are 24 single-qubit Clifford gates, including I , X , Y , Z , H , HX , HY , HZ , S , SX , SY , SZ , and others. Because of their properties under conjugation, Clifford group gates simplify the process of quantum gate teleportation.

Beyond this, the Clifford group also plays a crucial role in error correction within quantum computing. The examples below illustrate how teleportation is streamlined

when the gate to be teleported belongs to the Clifford group.

X Gate Teleportation

If we choose $U = X$ as the teleportation gate, we find that the properties of the X gate allow us to simplify the correction gates to basic Pauli matrices. This is due to the transformation properties of the X gate: $UXU^\dagger = XXX^\dagger = X$ and $UZU^\dagger = XZX^\dagger = -Z$. In essence, these correction gates are the same as those used in standard quantum teleportation, as detailed in § 10.3.4.4, albeit with an additional global phase factor.

ij	Qubit State	Correction Gates
00	$X \psi\rangle$	I or None
01	$XX \psi\rangle$	X
10	$XZ \psi\rangle$	$-Z$
11	$XZX \psi\rangle$	$-XZ$

H Gate Teleportation

The Hadamard gate (H) is one of the single-qubit Clifford gates. With $U = H$, $UXU^\dagger = HXH^\dagger = Z$ and $UZU^\dagger = HZH^\dagger = X$. Therefore, the correction gates become simple Pauli matrices as in the following table.

ij	Qubit State	Correction Gates
00	$H \psi\rangle$	None
01	$HX \psi\rangle$	Z
10	$HZ \psi\rangle$	X
11	$HZX \psi\rangle$	ZX

S Gate Teleportation

The S gate, also known as the phase gate, is a member of the Clifford group. With $U = S$, $UXU^\dagger = SXS^\dagger = Y$ and $UZU^\dagger = SZS^\dagger = Z$. Therefore, the correction gates become simple Pauli matrices as in the following table.

ij	Qubit State	Correction Gates
00	$S \psi\rangle$	None
01	$SX \psi\rangle$	Y
10	$SZ \psi\rangle$	Z
11	$SZX \psi\rangle$	YZ

T Gate Teleportation

The T gate, also known as the $\pi/8$ gate, given by $T = \sqrt{S}$, is not a Clifford gate. It doesn't map Pauli matrices to Pauli matrices under conjugation. Therefore, in the case of $U = T$, the correction gates won't simplify to single Pauli gates. Instead, we need to use the full form TXT^\dagger , TZT^\dagger , and $TXZT^\dagger$. However, these compensations

can be simpler to implement than the T gate itself. This property forms the basis for constructing fault-tolerant T gates.

Exercise 10.5 Find the correction gates for the teleportation of the Pauli Y gate and Z gate.

10.5.2 *Teleportation of Two-Qubit Gates

The gate teleportation process can be extended to include two-qubit unitary gates, such as the CNOT gate, or a combination of single and two-qubit gates, such as a Bell state generator. In this section, we explore two variations of two-qubit gate teleportation schemes: teleportation of one two-qubit gate and one qubit, which we refer to as 2-1; and teleportation of one two-qubit gate and two qubits, which we refer to as 2-2.

1 2-1 Gate Teleportation

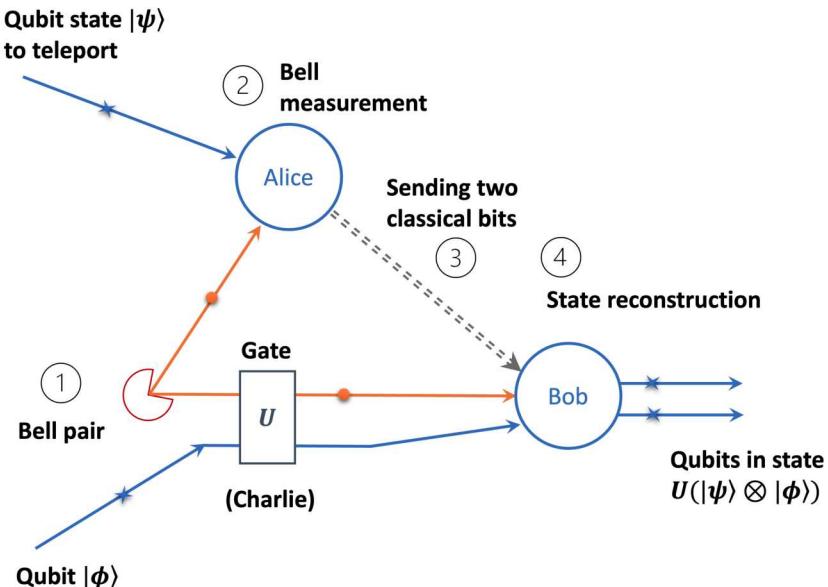


Figure 10.13: Basic Process of 2-1 Gate Teleportation

The 2-1 gate teleportation process is depicted in the diagram shown in Fig. 10.13 and in the quantum circuit in Fig. 10.14.

In this scenario, Alice has a qubit $|\psi\rangle$ that she wants to teleport to Bob. Charlie has a two-qubit gate U and another qubit $|\phi\rangle$. On Bob's end, we aim for him to have $U(|\psi\rangle \otimes |\phi\rangle)$ without directly accessing $|\psi\rangle$ or U .

To achieve this, Charlie prepares a Bell pair with qubits A and B . He sends qubit A to Alice. Alice performs a Bell measurement on $|\psi\rangle$ and qubit A , and sends the measurement results (two classical bits) to Bob via classical communication channel. Charlie also applies U to qubit B and $|\phi\rangle$, and sends both to Bob. Upon receiving the measurement results, Bob applies the appropriate correction gates to the qubits he received from Charlie, and successfully recovers $U(|\psi\rangle \otimes |\phi\rangle)$.

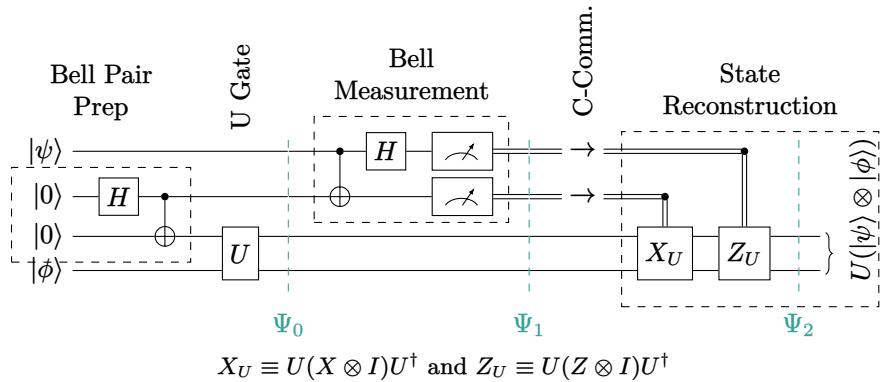


Figure 10.14: Quantum Circuit for 2-1 Gate Teleportation

1. Transition from Standard to Gate Teleportation

Let's now explore an intuitive connection between 2-1 gate teleportation and standard quantum teleportation, depicted in Fig. 10.15.

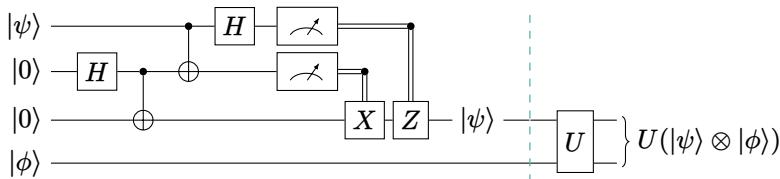


Figure 10.15: Gate U Applied Post-Teleportation

The segment prior to the dashed visual separator is the standard teleportation circuit from Fig. 10.5. At this circuit's output, the third qubit is in state $|\psi\rangle$. Upon application of a unitary gate U on both $|\psi\rangle$ and $|\phi\rangle$, the final state becomes $U(|\psi\rangle \otimes |\phi\rangle)$. This final state mirrors that of 2-1 gate teleportation in Fig. 10.14.

To transform the circuit in Fig. 10.15 into the gate teleportation circuit as shown in Fig. 10.14, we need to shift the gate U to the left of the controlled- Z and controlled- X gates. This transformation is facilitated by the following gate sequence equivalence, similar to the single-qubit gate teleportation scenario:

$$U(X \otimes I) = U(X \otimes I)(U^\dagger U) = (U(X \otimes I)U^\dagger)U = X_U U, \quad (10.35a)$$

$$U(Z \otimes I) = U(Z \otimes I)(U^\dagger U) = (U(Z \otimes I)U^\dagger)U = Z_U U, \quad (10.35b)$$

where $X_U \equiv U(X \otimes I)U^\dagger$ and $Z_U \equiv U(Z \otimes I)U^\dagger$. This explains the correction gates X_U and Z_U in Fig. 10.14.



Although quantum circuit diagrams apply gates from left to right, the corresponding operator multiplication operates in reverse order, from right to left.

2-1 Gate Teleportation: Core Summary

The initial state vector can be decomposed into the orthonormal Bell basis as follows:

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}} |\psi\rangle \otimes (I \otimes U)((|00\rangle + |11\rangle) \otimes |\phi\rangle) \quad (10.36a)$$

$$= \frac{1}{\sqrt{2}} (\alpha |0\rangle + \beta |1\rangle) \otimes (I \otimes U)((|00\rangle + |11\rangle) \otimes |\phi\rangle) \quad (10.36b)$$

$$= \frac{1}{2} \sum_{i,j \in \{0,1\}} |\beta_{ij}\rangle \otimes U(|\psi_{ij}\rangle \otimes |\phi\rangle), \quad (10.36c)$$

where $|\beta_{ij}\rangle$ represents the Bell states, and $|\psi_{ij}\rangle$ is defined in Table 10.2 and Eq. 10.22a.

When Alice measures the first two qubits in the Bell basis $\{\beta_{ij}\}$, each of the Bell states occurs with equal probability. If she obtains $|\beta_{ij}\rangle$ (where $i, j \in \{0, 1\}$), the state of the system collapses to $|\beta_{ij}\rangle \otimes U(|\psi_{ij}\rangle \otimes |\phi\rangle)$, and the third and fourth qubits take on the state $U(|\psi_{ij}\rangle \otimes |\phi\rangle)$. It can then be adjusted based on Alice's measurement results (i, j) to reconstruct the state $U(|\psi\rangle \otimes |\phi\rangle)$.

2. CNOT Gate Teleportation

You may question the necessity of teleporting a gate U when the correction gates, such as $U(X \otimes I)U^\dagger$, already contain U . It might appear more logical to teleport $|\psi\rangle$ first and then apply U directly. However, the procedure may be simplified when U is a specific two-qubit gate. In this section, we'll consider the CNOT gate as U .

For $U = \text{CNOT}$, it can be demonstrated through matrix multiplication (or through logic operation analysis - see next section) that

$$X_U \equiv U(X \otimes I)U^\dagger = X \otimes X, \quad (10.37a)$$

$$Z_U \equiv U(Z \otimes I)U^\dagger = Z \otimes I. \quad (10.37b)$$

Therefore, for $U = \text{CNOT}$, Fig. 10.14 simplifies to Fig. 10.16, in which the correction gates do not involve CNOT, achieving its true teleportation.

Exercise 10.6 Verify Eq. 10.37 for $U = \text{CNOT}$ through multiplication of the matrices of CNOT, X , and Z .

Exercise 10.7 Following the procedure for the derivation of the CNOT teleportation circuit in Fig. 10.16, design a practical circuit for the teleportation of CNOT with the last qubit (i.e., $|\phi\rangle$) as its control.

2-2 Gate Teleportation

Now we examine the teleportation of two qubits, $|\psi\rangle$ and $|\phi\rangle$, along with a two-qubit gate U . This operation is depicted in the diagram shown in Fig. 10.17 and quantum

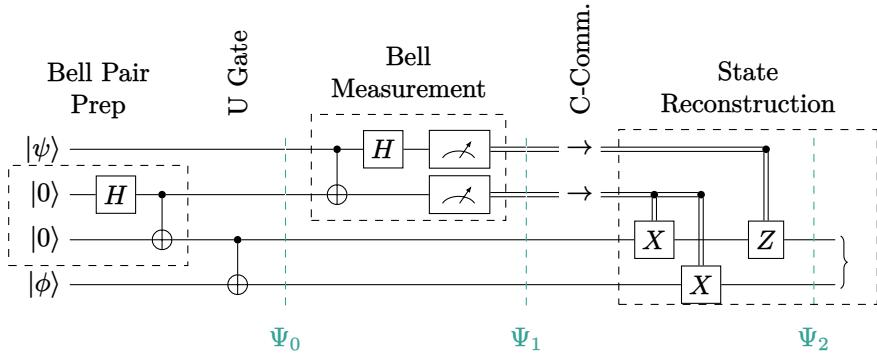


Figure 10.16: Quantum Circuit for 2-1 CNOT Teleportation

circuit in Fig. 10.18. In this scenario, Daniel prepares two Bell pairs: Pair *A* consists of qubits a_1 and a_2 , and Pair *B* consists of b_1 and b_2 . He sends a_1 from Pair *A* to Alice, and b_2 from Pair *B* to Bob. He then applies a two-qubit gate *U* to a_2 and b_1 , and sends them to Charlie.

Alice and Bob each possess an additional qubit, c_1 and c_2 , respectively, each in states $|\psi\rangle$ and $|\phi\rangle$. Their goal is to teleport these qubits while applying the *U* gate to them.

To accomplish this, Alice performs a Bell measurement on qubits c_1 and a_1 , while Bob does the same for c_2 and b_2 . They then transmit their respective measurement results to Charlie via a classical communication channel.

Upon receiving the measurement results, Charlie applies the appropriate correction gates to qubits a_2 and b_1 .

At the end of this process, Charlie is left with qubits a_2 and b_1 in the state $U|\psi\rangle|\phi\rangle$. This indicates that the states $|\psi\rangle$ and $|\phi\rangle$ have been teleported with the *U* gate applied to them. This sequence of operations effectively teleports the application of a *U* gate between two qubits originating from different locations.

1. Transition from Standard Teleportation

Similar to the illustration of transition from standard teleportation to single-qubit gate teleportation in § 10.5.1.1, let's do the same for two-qubit gate teleportation.

In Fig. 10.19, the segment prior to the dashed visual separator are two standard teleportation circuits in parallel, from Fig. 10.5. At the output, the qubits a_2 and b_1 are in the states $|\psi\rangle$ and $|\phi\rangle$, respectively. Upon application of a *U* gate, the qubit's state becomes $U|\psi\rangle|\phi\rangle$. This final state mirrors that of intended gate teleportation, wherein a *U* gate is applied to the teleported $|\psi\rangle$ and $|\phi\rangle$ (as depicted in Fig. 10.18).

To transform the circuit in Fig. 10.19 into the gate teleportation circuit as shown in Fig. 10.18, we need to shift the gate *U* to the left of the controlled-*Z* and controlled-*X* gates. This transformation is facilitated by the concept of gate sequence equivalence, which is demonstrated in the following table:

Here $X_1 \equiv X \otimes I$, $Z_1 \equiv Z \otimes I$, $X_2 \equiv I \otimes X$, and $Z_2 \equiv I \otimes Z$. (For example, X_2 represents the operator of an *X* gate acting on the second qubit.) This concept of gate

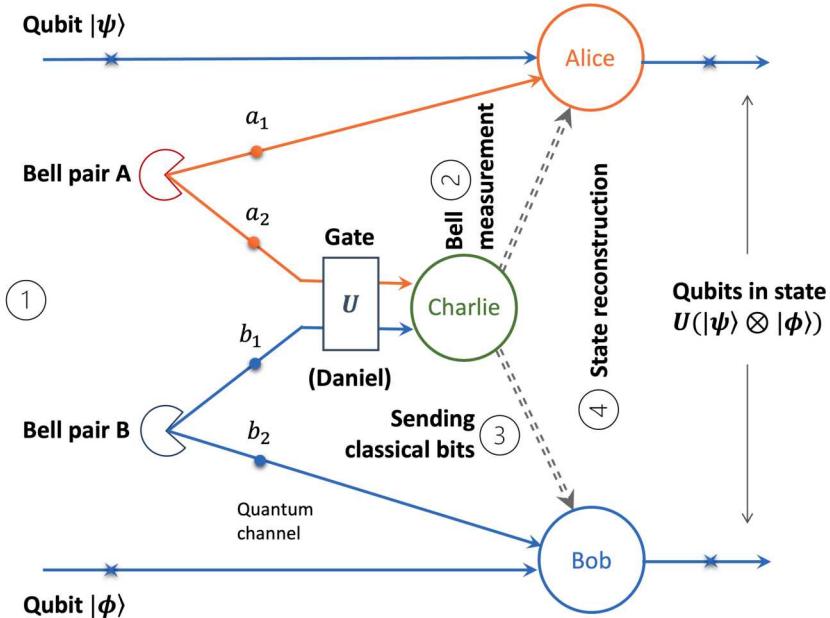


Figure 10.17: Basic Process of 2-2 Gate Teleportation

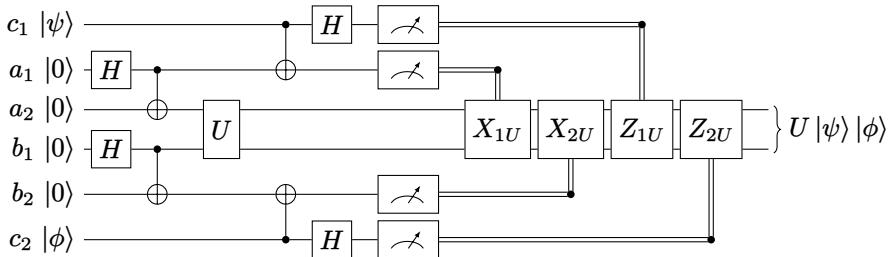


Figure 10.18: Quantum Circuit for 2-2 Gate Teleportation

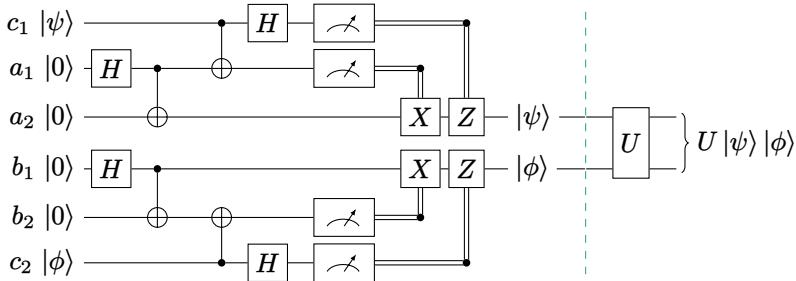
sequence equivalence underlies the definition of the correction gates $X_{1U} \equiv UX_1U^\dagger$, $Z_{1U} \equiv UZ_1U^\dagger$, $X_{2U} \equiv UX_2U^\dagger$, and $Z_{2U} \equiv UZ_2U^\dagger$ in Fig. 10.18.

2. CNOT Gate Teleportation

For the CNOT gate, the correction gates can be simplified. We can take advantage of the gate sequence equivalence shown in Table 10.6. For a detailed discussion of these sequences, see § 7.4.3. As you'll notice, in this scenario, the U correction gates are simplified to an additional X gate and Z gate, beyond the original four.

With the equivalent gate sequences in Table 10.6, we have finally arrived at the quantum circuit for CNOT gate teleportation in Fig. 10.20. The two extra gates due to gate sequence equivalence 2 and 3 are highlighted with shading. This is a highly practical circuit as the correction gates no longer involve U directly.

Exercise 10.8 Following the procedure for the derivation of the CNOT teleporta-

**Figure 10.19:** Two-Qubit U -Gate Applied Post-Teleportation

Gate Sequence	Equivalent Seq.	Operator Equivalence
$X \xrightarrow{} U$	$U \xrightarrow{} X_{1U}$	$UX_1 = UX_1(U^\dagger U) = (UX_1 U^\dagger)U = X_{1U}U$
$Z \xrightarrow{} U$	$U \xrightarrow{} Z_{1U}$	$UZ_1 = UZ_1(U^\dagger U) = (UZ_1 U^\dagger)U = Z_{1U}U$
$X \xrightarrow{} U$	$U \xrightarrow{} X_{2U}$	$UX_2 = UX_2(U^\dagger U) = (UX_2 U^\dagger)U = X_{2U}U$
$Z \xrightarrow{} U$	$U \xrightarrow{} Z_{2U}$	$UZ_2 = UZ_2(U^\dagger U) = (UZ_2 U^\dagger)U = Z_{2U}U$

Table 10.5: Gate Sequence Equivalence for Two-Qubit Gate Teleportation

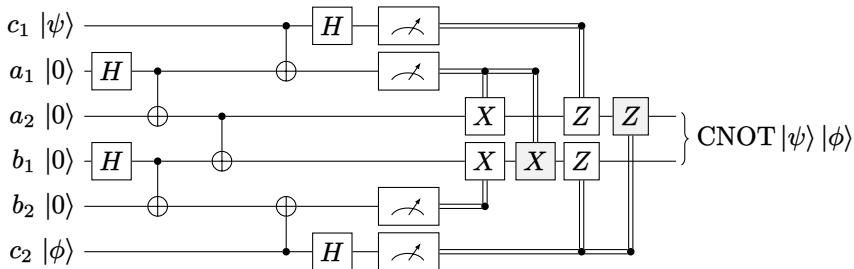
tion circuit in Fig. 10.20, design a practical circuit for the teleportation of the CZ gate.

10.5.3 *Further Exploration

The concept of quantum gate teleportation extends beyond the basic protocol discussed in this text. For those seeking to deepen their understanding, the following topics are suggested for further exploration:

1. Generalized Quantum Gate Teleportation [46]: This work introduces a method for creating a variety of quantum gates by teleporting qubits through specific entangled states. It delves deep into the theoretical framework and outlines potential applications, notably emphasizing a quantum computer model that employs single qubit operations, Bell measurements, and GHZ states.
2. Experimental Teleportation of a Quantum Controlled-NOT Gate [56]: This paper presents an empirical study where a CNOT gate is teleported. The experimental approach and the resulting observations offer valuable insights into the practical implications and challenges of gate teleportation.
3. Optimal Scheme for Teleportation of an n-qubit Quantum State [82]: This research introduces a quantum circuit tailored for teleporting an n -qubit quantum state. The scheme is noted for its optimization in terms of quantum resource utilization.

Index	Gate Sequence	Equivalent Sequence
1	$ x\rangle$ ————— $y\rangle$ ————— X ————— \oplus } $ \Psi_1\rangle$	$ x\rangle$ ————— $y\rangle$ ————— \oplus } X ————— } $ \Psi_2\rangle$
2	$ x\rangle$ ————— X ————— \oplus } $ \Psi_1\rangle$	$ x\rangle$ ————— $y\rangle$ ————— \oplus } X ————— } $ \Psi_2\rangle$
3	$ x\rangle$ ————— $y\rangle$ ————— Z ————— \oplus } $ \Psi_1\rangle$	$ x\rangle$ ————— $y\rangle$ ————— \oplus } Z ————— } $ \Psi_2\rangle$
4	$ x\rangle$ ————— Z ————— \oplus } $ \Psi_1\rangle$	$ x\rangle$ ————— $y\rangle$ ————— \oplus } Z ————— } $ \Psi_2\rangle$

Table 10.6: Gate Sequence Equivalence for CNOT Teleportation**Figure 10.20:** Quantum Circuit for CNOT Gate Teleportation

4. Teleportation of a Quantum Gate Between Two Logical Qubits [36]: A seminal piece that experimentally demonstrates the teleportation of a CNOT gate between two logical qubits. The research employs real-time adaptive control, making the teleportation deterministic. Additionally, it underscores the significance of error-correctable encoding for maintaining high process fidelity.
5. Fault-Tolerant Universal Quantum Gate Operations [74]: This paper provides a detailed account of fault-tolerant universal gate operations on logical qubits within a trapped-ion quantum computer. The research showcases the advantages of the “flag fault tolerance” paradigm and culminates in a demonstration of a fault-tolerant logical T gate. The findings pave the way for further advancements in error-corrected universal quantum computation.
6. Enhanced Quantum Teleportation with Logical States [60]: This recent paper proposes a novel approach to quantum teleportation, focusing on the transmission of multi-qubit physical states into error correctable multi-qubit logical states. The work also touches upon an efficient quantum error correction scheme, highlighting its potential in future fault-tolerant quantum technologies.

10.6 E91 Quantum Key Distribution Protocol

In cryptography, encryption entails the transformation of information (plaintext)

into a format only accessible by intended recipients through a shared secret key. The security hinges on the secure distribution of this key. Conventionally, transmitting messages through classical channels leaves them vulnerable to undetected eavesdropping. Quantum mechanics offers an alternative, providing the ability to transmit messages while detecting any eavesdropping.

In 1984, Charles H. Bennett and Gilles Brassard introduced the BB84 protocol which utilized four quantum states across two bases. Here, Alice sent particles in one of the four states, and Bob randomly selected bases to measure them. (The BB84 protocol is detailed in § 5.5.)

Building on BB84, Artur Ekert proposed the E91 Quantum Key Distribution (QKD) protocol in 1991. This protocol employs quantum entanglement, local operations, and classical communication channels to securely exchange cryptographic keys, with applications in secure communication and quantum-secured networks. A distinctive feature of the E91 QKD protocol is its intrinsic security, as any attempt to intercept the entangled photons alters their quantum states, leading to detectable anomalies in the shared key. Consequently, Alice and Bob can identify and discard compromised keys due to eavesdropping.

10.6.1 Procedure of the E91 QKD Protocol

The E91 QKD protocol encompasses five primary steps, as depicted in Fig. 10.21. We will explain these steps in detail next.

1 Bell Pair Generation and Distribution

A Bell pair source generates pairs of entangled qubits, typically in the Bell state $|\Psi^-\rangle$. One qubit from each pair is sent to Alice, and the other to Bob.

In the E91 protocol, unlike BB84 where Alice sends qubits to Bob, a central source creates and distributes the entangled qubits. This configuration is considered more practical for real-world implementations, especially over long distances, where a central source such as a satellite can distribute entangled qubits to multiple receivers.

2 Measurement

Upon receiving the qubits, Alice and Bob independently and randomly choose one of three predetermined measurement bases to measure their respective qubits.

The three bases for Alice are:

- a_1 corresponds to measuring in the Z basis (0° on the Bloch sphere).
- a_2 corresponds to measuring in a basis that is 45° rotated from the Z basis on the Bloch sphere, often denoted as $W = \frac{1}{\sqrt{2}}(Z + X)$.
- a_3 corresponds to measuring in the X basis (90° on the Bloch sphere).

For Bob:

- b_1 corresponds to measuring in the W basis (45° on the Bloch sphere).
- b_2 corresponds to measuring in the X basis (90° on the Bloch sphere).
- b_3 corresponds to measuring in a basis that is 135° rotated from the Z basis on the Bloch sphere, often denoted as $V = \frac{1}{\sqrt{2}}(Z - X)$.

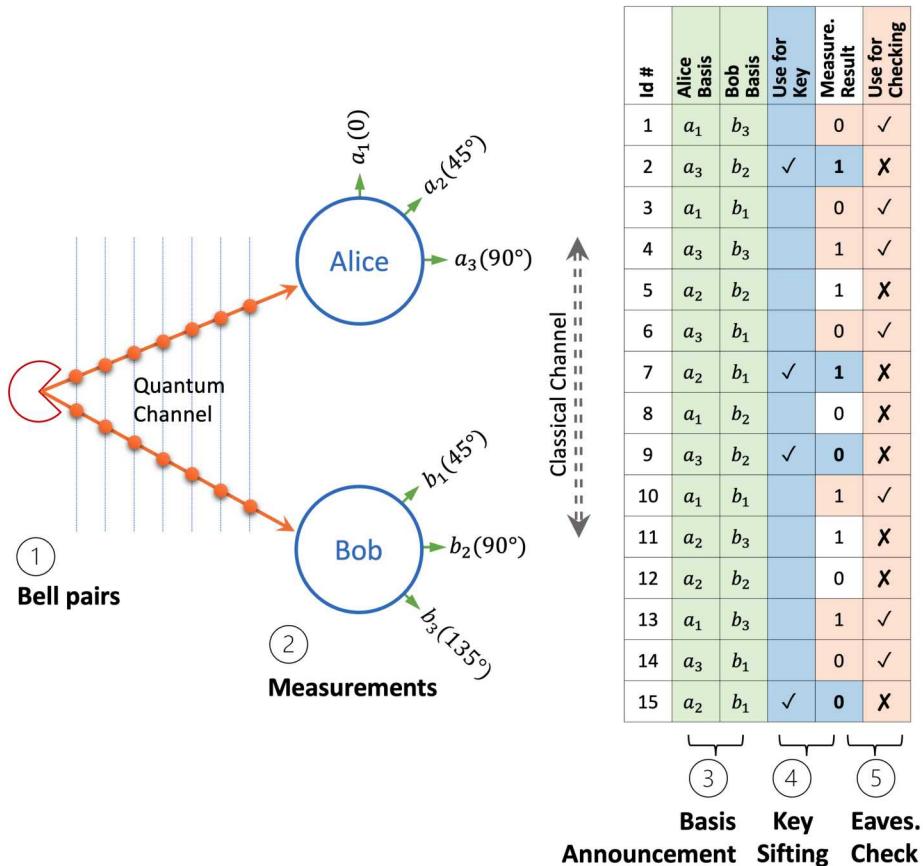


Figure 10.21: Basic E91 QKD Protocol

The BB84 QKD protocol utilizes four measurement settings, two for each participant, while the E91 protocol uses six, with three for each participant. The additional settings in E91 facilitate eavesdropper detection through Bell inequality tests, minimizing the need to reveal key information during the process. This will become clear after we examine the subsequent steps.



For spin-1/2 systems, the angles on the Bloch sphere equal to the angles of the physical orientation of the Stern-Gerlach apparatus. For photon polarization, however, the physical angles of the polarizers are half the angles on the Bloch Sphere, that is, for Alice a_1 corresponds to 0° , a_2 22.5° , and a_3 45° ; and for Bob, b_1 22.5° , b_2 45° , and b_3 67.5° .

3 Basis Announcement

After the measurements, Alice and Bob announce the bases they used for each measurement over a public classical channel. They do not disclose their measurement results at this stage.

4 Key Sifting

Using the announced bases, Alice and Bob independently sift their data to retain only the measurement results performed in pairs of compatible bases (a_2, b_1 and a_3, b_2). For these pairs, the measurement results of Alice and Bob are anticorrelated when measuring an entangled state; i.e., if Alice obtains a result of 1, Bob obtains a result of 0, and vice versa (one of them must invert their measurement results).

They retain the results measured with compatible bases as a sifted key. These are indicated in the “Used for Key” column in the table of Fig. 10.21.

5 Eavesdropping Check

Alice and Bob use a subset of their measurement results to perform a Bell’s inequality test, designed to detect potential eavesdropping.

The subset includes results from bases a_1, a_3, b_1 , and b_3 , i.e., excluding the results from a_2 and b_2 . These are indicated in the “Used for Checking” column in the table of Fig. 10.21.

The specific choice of angles for the bases a_1, a_3, b_1 , and b_3 is designed to maximize the value of the Bell CHSH quantity, which reaches $|S| = 2\sqrt{2}$ for maximally entangled states. Normally, without eavesdropping, the measured photons are in a Bell state and $|S| = 2\sqrt{2}$. However, if an eavesdropper has tampered with the qubits, then $|S| < 2\sqrt{2}$.

By comparing a portion of their measurement results (announced over a classical channel), Alice and Bob can identify any violations of Bell’s inequality. If the violations exceed the quantum limit, they conclude that eavesdropping has occurred, and the generated key is unsafe.

6 Key Generation

If the channel is deemed secure following the eavesdropping check, Alice and Bob use the remaining sifted key data from the compatible bases for secure communication. They can also apply error correction and privacy amplification techniques to enhance the security of this key.

10.6.2 Comparison to the BB84 Protocol

Both the E91 and BB84 protocols are Quantum Key Distribution (QKD) protocols that aim to establish a secure key between two parties, typically referred to as Alice and Bob. Despite their shared goal, the two protocols differ significantly in their mechanisms and practical implications. (We have explored the BB84 protocol in § 5.5.)

1 Basis of Security

- The E91 protocol is rooted in quantum entanglement and Bell’s inequalities. Alice and Bob each receive one of an entangled pair of qubits. The quantum correlations between measurements on these entangled pairs are exploited to detect the presence of an eavesdropper.
- The BB84 protocol is based on Heisenberg’s Uncertainty Principle. An eavesdropper (Eve) cannot measure the quantum states sent by Alice without

causing a disturbance. This disturbance can be detected by Alice and Bob during the error-checking phase.

2 Detection of Eavesdropping

- In E91, eavesdropping is detected by a decrease in the correlation between Alice's and Bob's measurement outcomes, indicating a violation of Bell's inequalities. This can be observed without any further communication between Alice and Bob beyond sharing their measurement results.
- In BB84, Alice and Bob detect eavesdropping by comparing a subset of their key bits. If Eve tries to measure the qubits, she introduces errors because of the uncertainty principle, which Alice and Bob can detect.

3 Number of Bases Used

- The E91 protocol uses three bases for encoding and measuring qubits. Alice and Bob randomly select one of the three bases to measure their qubits for each run.
- The BB84 protocol utilizes two non-commuting bases for encoding and measuring qubits. Alice randomly selects one of these bases to encode each bit of the key, and Bob randomly selects one of the bases to measure each bit.

4 Practicality for Real-world Applications

- The E91 protocol, with its reliance on entangled pairs of qubits, can be technologically challenging to implement, especially in terms of creating, maintaining, and transmitting entangled qubits over long distances. Additionally, the requirement to measure in three different bases may add to the complexity. However, in scenarios where a trusted source can distribute entangled photons, such as satellite-based quantum communication, E91 can have advantages in terms of security and efficiency for long-distance quantum key distribution.
- The BB84 protocol is often considered more practical for short-distance real-world applications as it requires only the preparation and transmission of single qubits and the capability to measure in two bases. These tasks are relatively easier to achieve with current technology, making BB84 more straightforward for implementation. However, the protocol necessitates a more interactive communication process between Alice and Bob to detect eavesdropping, which can be a limitation in certain scenarios.

10.6.3 *Further Exploration

The concept of the E91 Quantum Key Distribution Protocol, as well as its basis in entanglement and the broader context of quantum cryptography, extends beyond the basic protocol discussed in this text. For those seeking to deepen their understanding of the E91 protocol, its foundations, and its comparisons with other quantum cryptographic schemes, the following are suggested for further exploration:

1. Quantum Cryptography Based on Bell's Theorem [39]: This is the original paper by Artur Ekert, in which he introduced the E91 protocol. It lays the foundation for Quantum Key Distribution based on entanglement and Bell's inequalities.

2. Quantum Key Distribution Network for Multiple Applications [78]: This paper discusses the implementation of a quantum key distribution network where different protocols, including E91, are employed for various applications. It demonstrates the flexibility and real-world utility of the E91 protocol in a network setting.
3. Experimental Demonstration of a Measurement-Device-Independent Quantum Key Distribution [76]: This paper introduces and demonstrates an experimental implementation of measurement-device-independent quantum key distribution, which can be seen as an advancement over the E91 protocol, by removing the need to trust the measurement devices.
4. Quantum Cryptography: Public Key Distribution and Coin Tossing [21]: Though primarily centered around the introduction of the BB84 protocol, this seminal paper by Charles H. Bennett and Gilles Brassard also offers a basis for understanding the context in which the E91 protocol was developed.
5. Quantum Key Distribution in Cryptography: A Survey [13]: This article explores the integration of quantum key distribution (QKD) in cryptographic infrastructures. Highlighting QKD's promise of information-theoretic security, the review underscores its application in renewing symmetric cipher keys and enabling secure key establishment in networks, while addressing inherent challenges and research avenues.

10.7 Summary and Conclusions

Entanglement in Applications

This chapter has taken readers on a comprehensive exploration of the myriad applications of quantum entanglement in quantum computing and quantum communication. Beginning with an overview of the essential role entanglement plays in these domains, we ventured into a systematic analysis of some of the key protocols that harness the power of entanglement.

Our journey began with the elucidation of superdense coding, an elegant method that demonstrated the efficacy of quantum communication. Transitioning from communication to computation, the concept of quantum teleportation showcased the unique capability of transmitting quantum states without the actual physical transfer of qubits, highlighting the distinction between classical and quantum paradigms.

Entanglement swapping, a nuanced yet fundamental process, further underscored the versatility of entanglement in establishing quantum connections over long distances. Such a technique is paramount in the construction of quantum networks, which will potentially revolutionize global communication in the quantum era.

We delved deeper into computation with the discussion of quantum gate teleportation. This approach opens doors to distributed quantum computing, heralding a paradigm where quantum gates can be operationalized across spatially dispersed quantum systems.

Finally, we transitioned back into the realm of quantum communication with the E91 Quantum Key Distribution (QKD) protocol. Rooted in entanglement and the principles of quantum mechanics, this protocol emphasizes the enhanced security

quantum systems offer in the face of eavesdropping threats, distinguishing quantum cryptographic techniques from their classical counterparts.

Bridging Theory and Practice

Throughout this chapter, our approach has been two-pronged: we have grounded our discussions in the theoretical constructs of quantum mechanics while constantly alluding to their real-world implications. By doing so, we intend to foster an appreciation for both the abstract beauty of quantum phenomena and their tangible applications in next-generation technologies.

Entanglement, as a quintessential quantum resource, stands at the confluence of many quantum applications. Through the varied techniques and protocols discussed, we aim to have imparted a nuanced understanding of its potential and the avenues it opens up in quantum information science.

Upcoming Topics

Our subsequent chapter promises to immerse readers into the intricate world of quantum algorithms. We will juxtapose quantum algorithms against their classical counterparts, shedding light on the unique advantages and challenges they present. Starting with the pedagogical Deutch Jozsa algorithm, we will then transition into the realm of optimization with algorithms such as CUBO, QAOA, and VQE, which have immediate real-world applications. We will conclude the chapter with a dive into the Quantum Measurement Bomb, an algorithm that underscores the central role of measurements in quantum computing. As we venture into these topics, we will continually build on the foundational principles expounded in the previous chapters, reinforcing and expanding your quantum computing knowledge.

Problem Set 10

10.1 Investigate the following questions related to **superdense coding** and elaborate your answers:

- (a) Why is superdense coding considered more efficient than classical transmission if both methods involve two bits?
- (b) Does superdense coding enhance transmission security?
- (c) How is superdense coding related to quantum teleportation?
- (d) Can a single EPR pair be used to transmit more than two bits of classical information?

10.2 Investigate the following questions related to **quantum teleportation** and elaborate your answers:

- (a) Can quantum teleportation teleport matter, i.e., move particles from one location to another?
- (b) Is quantum teleportation faster than light?

- (c) Can quantum teleportation be used to copy quantum states?
- (d) After Alice measures the two qubits, are the original qubit pair still entangled?
- (e) Can the entangled pair be used twice for quantum teleportation?
- (f) What is the relationship between quantum teleportation and the no-cloning theorem?
- (g) What is the role of classical communication in quantum teleportation?

10.3 Investigate the following questions related to **entanglement swapping** and elaborate your answers:

- (a) How is entanglement swapping related to quantum teleportation?
- (b) Is entanglement swapping instantaneous?
- (c) Is it possible to swap entanglement more than once?
- (d) Why is a mediator required in entanglement swapping?
- (e) Can entanglement swapping be used to transmit classical information?
- (f) How is entanglement swapping used in a quantum network?
- (g) How does the measurement result of the mediator influence the final state of the other two qubits?
- (h) Is there a risk of losing entanglement during the swapping process?
- (i) Can the mediator control the type of entanglement that is created between the two remaining qubits?

10.4 Investigate the following questions related to **quantum gate teleportation** and elaborate your answers:

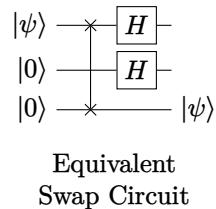
- (a) How does quantum gate teleportation relate to regular quantum teleportation?
- (b) What is the practical use of quantum gate teleportation?
- (c) How does quantum gate teleportation impact quantum computing?
- (d) How does quantum gate teleportation contribute to quantum cryptography?
- (e) What are the limitations of quantum gate teleportation?
- (f) Can quantum gate teleportation be achieved without classical communication?
- (g) How far apart can two qubits be for successful quantum gate teleportation?
- (h) What does it mean to say that the states $|\psi\rangle$ and $|\phi\rangle$ have been teleported with the U gate applied to them?

10.5 Investigate the following questions related to **the E91 QKD protocol** and elaborate your answers:

- (a) What advantages does the use of entanglement in E91 provide over the BB84 protocol?

- (b) Can the E91 protocol be used for communication beyond key distribution?
- (c) How does the E91 protocol ensure that the key has not been compromised?
- (d) What are the challenges in implementing the E91 protocol in real-world scenarios?
- (e) Can the E91 protocol work with different types of particles, or is it restricted to photons?
- (f) How does the E91 protocol deal with noise and loss in a real-world communication channel?
- (g) What is the fraction of measurement bits that can be used as sifted keys?
- (h) Can the E91 protocol be used in a network with more than two parties?
- (i) Is the E91 protocol widely adopted in commercial quantum communication systems?

- 10.6** Upon reflection, you'll realize that the end result of quantum teleportation is essentially swapping the first and third qubits. In fact, Fig. 10.6 (excluding the final measurements) is equivalent to the swap circuit shown on the right. Your task is to demonstrate this equivalence using gate sequence equivalence developed in §§ 7.4 and 8.5.



- 10.7** Consider the scenario in quantum teleportation where Alice and Bob share the Bell state $|\Psi^-\rangle$ (commonly associated with entangled photon pairs), as opposed to $|\Phi^+\rangle$. The objective is to work through the quantum teleportation protocol given this condition. Specifically, you are tasked with deriving an equation analogous to Eq. 10.23, which concisely encapsulates the principles of quantum teleportation under the premise of the shared $|\Psi^-\rangle$ state.

- 10.8** Alice has the 2-qubit state $\alpha|00\rangle + \beta|11\rangle$. Demonstrate the process of quantum teleportation by which Alice can send this state to Bob. Specifically, detail the steps she must take to teleport the state using only 2 classical bits of communication and one shared EPR pair. Assume the EPR pair is in the state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and is distributed between Alice (first qubit) and Bob (second qubit) before teleportation begins. Outline how measurement outcomes affect the protocol and how Bob can reconstruct the original state.

- 10.9** Explore the entanglement swapping protocol assuming that the EPR pairs are in the Bell state $|\Psi^-\rangle$, a state often encountered in entangled photon pairs. Specifically, derive a formulation analogous to Eq. 10.30 that accurately describes the entanglement swapping process when the initial state of the entangled pairs is $|\Psi^-\rangle$.

- 10.10** Gate teleportation is a sophisticated protocol that extends the concept of quantum teleportation by incorporating quantum gates into the teleportation process. This allows for the execution of quantum gates on a quantum state remotely, without directly interacting with the state itself. In this exercise, consider a scenario where the entangled resource shared between the parties (Alice and Bob) is the Bell state $|\Psi^-\rangle$. Your objective is to work through the

gate teleportation protocol given this specific initial entangled state. Derive a formulation equivalent to Eq. 10.34 that succinctly represents the gate teleportation process utilizing the $|\Psi^-\rangle$ state.

- 10.11** Following the procedure for the derivation of the **2-1** CNOT teleportation circuit in Fig. 10.16, design a practical circuit for the teleportation of the Bell state generator, i.e., a Hadamard gate followed by a CNOT gate on the control qubit.
- 10.12** Following the procedure for the derivation of the **2-2** CNOT teleportation circuit in Fig. 10.20, design a practical circuit for the teleportation of the Bell analyzer, i.e., CNOT followed by a Hadamard gate on the control qubit.
- 10.13** * Consider the multipartite Bell states of four qubits as defined in Problem 8.8. These extended Bell states offer a rich framework for advancing quantum communication protocols beyond two-qubit entanglement. Your task is to investigate and extend the following core quantum communication protocols using these states:
- Explore how the multipartite Bell states can be utilized to enhance the efficiency of superdense coding. Consider the implications of using four-qubit states for encoding and transmitting more information than traditional two-qubit Bell states.
 - Extend the concept of quantum teleportation to incorporate four-qubit Bell states.
 - Extend the concept of entanglement swapping to incorporate four-qubit Bell states.
 - Investigate the generalization of superdense coding, quantum teleportation, and entanglement swapping using extended Bell states of $2n$ qubits.

In each case, provide a theoretical foundation for your extensions, supported by mathematical formulations where applicable.



Quantum Computation & Information

11	Quantum Algorithms: A Sampler	288
12	Quantum Error Correction: A Primer	333
13	Fundamentals of Quantum Information ...	396

11. Quantum Algorithms: A Sampler

Contents

11.1	The Deutsch-Jozsa Algorithm	290
11.1.1	Problem Statement	290
11.1.2	Classical vs. Quantum	290
11.1.3	Procedure of the Deutsch-Jozsa Algorithm	291
11.1.4	* Generalized Deutsch-Josza Algorithm	293
11.1.5	* The Phase Oracle	296
11.2	QUBO, VQE, QAOA, and AQC	298
11.2.1	The Max-Cut Problem	300
11.2.2	* The Ising Model	303
11.2.3	* Variational Quantum Eigensolver (VQE)	304
11.2.4	* Quantum Approximate Optimization Algorithm (QAOA)	307
11.2.5	* Traveling Salesman Problem	310
11.2.6	Adiabatic Quantum Computation and Annealing	312
11.3	Quantum Bomb and Quantum Money	315
11.3.1	The Quantum Bomb Test Algorithm	315
11.3.2	* Basics of Quantum Money \$	318
11.4	Summary and Conclusions	328
	Problem Set 11	329

An algorithm generally refers to a defined set of instructions or procedures that a computer adheres to in order to solve a particular problem. When these algorithms are designed to run on quantum computers, they are termed quantum algorithms. The objective behind quantum algorithms often centers around executing calculations more efficiently than classical computers, or addressing computations that are practically infeasible for classical computers due to inherent physical limitations (e.g., execution time or memory capacity). In this text, we classify quantum algorithms into three primary categories: Foundational Quantum Algorithms, NISQ Hybrid Algorithms, and Innovative Quantum Algorithmic Concepts.

Quantum Utility Era

As of circa 2023, quantum computing is primarily operating with Noisy Intermediate-Scale Quantum (NISQ) devices. These devices, though limited in scale and afflicted by quantum noise, represent a critical transition phase. They possess computational capabilities that begin to exceed those achievable by classical computer simulations.

In this evolving landscape, we are entering the quantum utility era, a phase where quantum computing starts to yield substantial computational advantages for particular tasks. This marks the onset of quantum computing's practical impact, as it begins to offer tangible solutions to real-world problems, despite the inherent challenges of quantum noise and device limitations.

Evolving Hybrid Algorithms

As quantum computing transitions from the NISQ era to the emerging era of quantum utility, quantum-classical hybrid algorithms are evolving to harness the strengths of both quantum and classical computing paradigms. These algorithms are pivotal for optimizing the performance of both NISQ devices, with their limited qubits and susceptibility to errors, and more advanced utility-scale quantum systems. By strategically integrating quantum computations with classical processes, these hybrid algorithms aim to enhance computational reliability and efficiency, adapting to the evolving capabilities of quantum hardware.

Foundational Quantum Algorithms

Historically, quantum algorithms that showcased the fundamental advantages of quantum over classical computation, such as Shor's and Grover's algorithms, were developed with the assumption of large-scale, fault-tolerant quantum computers for their operation. Such algorithms, which exploit the superposition and entanglement properties of qubits, are termed Foundational Quantum Algorithms. These algorithms serve as pivotal pedagogical tools and epitomize the profound possibilities of quantum computation.

Innovative Quantum Algorithmic Concepts

Beyond the established paradigms of NISQ and foundational algorithms lies a category of innovative quantum algorithmic concepts. These include algorithms and ideas that may have been conceived in the early stages of quantum computing or represent novel approaches in the field. They extend the boundaries of quantum computational research, exploring new methodologies and challenging existing assumptions. Among these are the Quantum Bomb Test Algorithm and Quantum Money, which offer insights into quantum measurements, probabilistic algorithms, and quantum cryptography.

Our Game Plan

In this chapter, we will explore representative quantum algorithms from each of the aforementioned categories:

Deutsch Jozsa Algorithm: A foundational quantum algorithm, it showcases the potential advantages of quantum computation. While its tangible real-world applications may be elusive, we will examine the intuition, working principles, and the generalization of this cornerstone algorithm.

CUBO, QAOA, VQE, AQC: As hallmarks of the NISQ era, these algorithms

have proven adept at addressing a myriad of optimization and quantum simulation challenges. Notably, this includes challenges solvable by quantum annealing devices. As a focal point, we will delve into the max-cut problem.

Quantum Measurement Bomb Algorithm and Quantum Money: Representing innovative quantum algorithmic concepts, these examples provide insights into various distinctive aspects such as quantum measurements, probabilistic algorithms, and quantum cryptography.

Through our exploration of these algorithms, we aim to provide readers with a comprehensive understanding of the evolving landscape of quantum computing, tracing its past achievements, present capabilities, and future potential.

11.1 The Deutsch-Jozsa Algorithm

The Deutsch-Jozsa (DJ) algorithm stands as one of the earliest and most elementary instances of a quantum algorithm. While it lacks direct practical applications, it exemplifies the potential advantage of quantum computing over classical computing in specific tasks.

11.1.1 Problem Statement

Consider a binary function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, which takes an n -bit string input and returns a single bit as output. The function is guaranteed to have one of two behaviors:

1. Constant: $f(x)$ is the same for all inputs; that is, either $f(x) = 0$ or $f(x) = 1$ for all x .
2. Balanced: For half of the possible inputs (not necessarily the first half), $f(x) = 0$, and for the other half, $f(x) = 1$.

The algorithm employs an *oracle*, a blackbox mechanism representing the function. One can query the oracle to understand the input-output relationship of the function, but the internal mechanics remain concealed. The objective of the algorithm is to determine with certainty whether the function is constant or balanced across its inputs.



In a theoretical context, the oracle is conceptualized as the black-box function input to the computation. The algorithm simply assumes its existence without concern for its internal structure. However, for pedagogical completeness, we provide a possible implementation of this oracle in § 11.1.5, which will aid readers in gaining a deeper understanding of the algorithm's operation.

Exercise 11.1 Enumerate all possible constant and balanced functions for $n = 1$ and $n = 2$.

11.1.2 Classical vs. Quantum

For a classical approach, the task of distinguishing whether the function $f(x)$ is constant or balanced necessitates a worst-case scenario of at least $2^{n-1} + 1$ oracle

queries.

In contrast, the DJ algorithm showcases the power of quantum computing by achieving this differentiation with just a single oracle query, presenting an exponential speedup in comparison to its classical counterpart. This disparity in performance underscores the profound advantage introduced by quantum parallelism.

11.1.3 Procedure of the Deutsch-Jozsa Algorithm

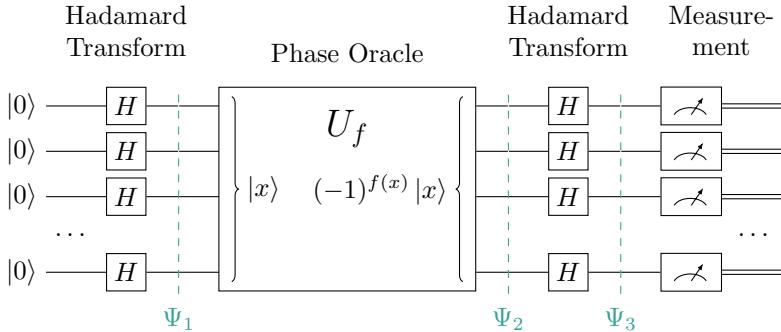


Figure 11.1: Quantum Circuit for the Deutsch-Jozsa Algorithm

As shown in Fig. 11.1, the DJ algorithm involves the following procedure:

1. Initialize all qubits to $|0\rangle$.
2. Apply a Hadamard transform to create a superposition of all possible input states.
3. Query the phase oracle, introducing a phase $(-1)^{f(x)}$ to each state in the superposition.
4. Execute the second Hadamard transform.
5. Measure the qubits. A constant function always produces the state $|000\dots 0\rangle$, while a balanced function results in any state except $|000\dots 0\rangle$.

We will elaborate on the procedure of the DJ algorithm next. While the DJ algorithm might not possess direct practical applications, the strategies and principles elucidated here lay foundation for understanding of more sophisticated quantum algorithms, notably Simon's, Shor's, and Grover's algorithms.

1 Uniform Superposition

Applying a Hadamard gate on each qubit initialized to $|0\rangle$ yields:

$$|\Psi_1\rangle = \bigotimes_{i=1}^n H|0\rangle = \bigotimes_{i=1}^n \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle). \quad (11.1)$$

Expanding the tensor product gives:

$$|\Psi_1\rangle \equiv |\Psi_u\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle. \quad (11.2)$$

Here $|\Psi_u\rangle$ is a uniform superposition of the 2^n basis states, from $|000\dots 0\rangle$ to $|111\dots 1\rangle$.

Exercise 11.2 Compute $\bigotimes_{x=1}^n H |1\rangle$. Is this also a uniform superposition of the computational basis?

2 The Phase Oracle

In the context of our quantum circuit, querying the function $f(x)$ necessitates its representation as a unitary transformation, capable of handling superposition states. An innovative realization of this unitary representation is termed the “phase oracle”, denoted by U_f . (A detailed account of the oracle’s implementation will be provided in § 11.1.5.)

Each potential input string for the function $f(x)$ maps to a distinct basis state, $|x\rangle$, within the n -qubit system. The primary role of the phase oracle is to impart a phase shift of π to a given $|x\rangle$ if its corresponding $f(x) = 1$:

$$U_f |x\rangle = (-1)^{f(x)} |x\rangle, \quad \text{for } x \in \{0, 1\}^n. \quad (11.3)$$

After applying the oracle, the state becomes:

$$|\Psi_2\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0, 1\}^n} (-1)^{f(x)} |x\rangle. \quad (11.4)$$

Exercise 11.3 Assume $f(x) = \text{parity}(x)$, and the input state of the oracle is $\frac{1}{\sqrt{2}}(|00\dots 0\rangle + i|11\dots 1\rangle)$. Compute the output state.

3 Hadamard Transform and Measurement

In typical scenarios, one would express the state $|\Psi_3\rangle$ as a result of applying the second Hadamard transform to $|\Psi_2\rangle$. However, within the context of the DJ algorithm, articulating this explicitly is unnecessary. Our primary focus lies in discerning the probability associated with the measurement outcome of $|000\dots 0\rangle$ in the state $|\Psi_3\rangle$.

The component $|000\dots 0\rangle$ in $|\Psi_3\rangle$ originates from the uniform superposition of all basis states in $|\Psi_2\rangle$, denoted as $|\Psi_u\rangle$ in Eq. 11.2.

Therefore, the probability of obtaining the outcome $|000\dots 0\rangle$ hinges on the overlap between states $|\Psi_u\rangle$ and $|\Psi_2\rangle$:

$$P_0 = |\langle \Psi_u | \Psi_2 \rangle|^2. \quad (11.5)$$

Upon inspection, we find that if $f(x)$ is a constant function, the state $|\Psi_2\rangle$ retains its uniform superposition, leading to $P_0 = 1$. Conversely, for a balanced function $f(x)$, half the terms in $|\Psi_2\rangle$ carry a negative sign. This results in perfect destructive interference and consequently, $P_0 = 0$. This distinction serves as the central insight behind the DJ algorithm.

11.1.4 *Generalized Deutsch-Josza Algorithm

In the traditional Deutsch-Josza (DJ) algorithm, measurement is performed against the $|000\dots0\rangle$ basis state to classify functions as either constant or balanced. Extending these measurements to encompass all computational basis states, however, provides deeper insights into the function's periodicity. While this generalization aligns with the principles of the Bernstein-Vazirani algorithm mathematically, it also enhances our understanding of how the DJ algorithm connects to the broader concept of the Fourier transform. This extension not only broadens the scope of problems that can be tackled but also underscores the fundamental relationship between quantum computing algorithms and Fourier analysis.

1 Building Intuition

Each computational basis state can be associated with a unique periodic function. For elucidation, consider the Hadamard transform H_4 , illustrated in Table 11.1. Here, the basis state $|0001\rangle$ is tied to a function with a period of 2. Conversely, states $|0010\rangle$ and $|0011\rangle$ indicate functions with a period of 4, with $|0011\rangle$ introducing a phase shift distinct from the square wave of $|0010\rangle$. Similarly, states from $|0100\rangle$ through $|0111\rangle$ denote functions with a period of 8. Furthermore, these functions are orthogonal to each other. In fact, they serve as the basis functions of the Hadamard transform, analogous to the pure harmonics in the Fourier transform.

Row	Basis	Period	Matrix Elements														
0	0000	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0001	2	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1
2	0010	4	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1
3	0011	4	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1
4	0100	8	1	1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1
5	0101	8	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	1	-1	1
6	0110	8	1	1	-1	-1	-1	1	1	1	1	-1	-1	-1	1	1	1
7	0111	8	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	-1
8	1000	16	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1
9	1001	16	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1
10	1010	16	1	1	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1
11	1011	16	1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	1	-1	1
12	1100	16	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1
13	1101	16	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1
14	1110	16	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1
15	1111	16	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	1

Table 11.1: Hadamard Transform H_4 and Period Functions

Exercise 11.4 For H_6 , determine the period of $f(x)$ corresponding to the basis state $|001100\rangle$. How many periodic occurrences does $f(x)$ exhibit over the 2^6 possible x values?

2 Problem and Solution

Quantum mechanics posits that orthogonal states can be distinctly identified through measurements. Inspired by this observation, we propose a more encompassing version

of the DJ algorithm:

Given a function in the basis function set of the 2^n -dimensional Hadamard transform, $f : \{0, 1\}^n \rightarrow \{0, 1\}$, mapping an n -bit string to a single-bit outcome, f is certain to display circular periodicity with periods spanning from 1 (for a constant function), through 2, 4, ..., 2^i , ..., 2^n . Our primary goal is to discern this periodicity.

While a classical method requires at least n function queries, quantum computing leverages the generalized DJ algorithm to accomplish this in merely one query.

The quantum methodology still relies on the original DJ circuit illustrated in Fig. 11.1. Given a measurement outcome of $|\tilde{y}\rangle$, the period is 2^{n-k} , where k denotes the count of leading 0s in \tilde{y} 's binary representation.

3 Proof

The Hadamard transform maps the computational basis state $|x\rangle$ into a superposition state of $|y\rangle$:

$$|x\rangle \xrightarrow{H^{\otimes n}} \frac{1}{\sqrt{2^n}} \sum_{y \in \{0,1\}^n} (-1)^{x \cdot y} |y\rangle, \quad (11.6)$$

where $x \cdot y$ represents the binary dot product of x and y , i.e., $x \cdot y = \bigoplus_{i=1}^n x_i y_i$, with $x_i, y_i \in \{0, 1\}$.

The multi-qubit state across different junctures of the DJ circuit, shown in Fig. 11.1, evolves as:

$$|0\rangle^n \xrightarrow{H^{\otimes n}} |\Psi_1\rangle \equiv |\Psi_u\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle \quad (11.7a)$$

$$\xrightarrow{U_f} |\Psi_2\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} |x\rangle \quad (11.7b)$$

$$\xrightarrow{H^{\otimes n}} |\Psi_3\rangle = \frac{1}{2^n} \sum_{x, y \in \{0,1\}^n} (-1)^{f(x) \oplus (x \cdot y)} |y\rangle. \quad (11.7c)$$

Exercise 11.5 Explain why $(-1)^{f(x)+x \cdot y} = (-1)^{f(x) \oplus (x \cdot y)}$, as utilized in the final transformation of Eq. 11.7.

A key insight of the extended DJ algorithm is that if $f(x)$ is

$$f(x) = x \cdot \tilde{y} \quad (11.8)$$

for a particular $\tilde{y} \in \{0, 1\}^n$, then $f(x) \oplus (x \cdot \tilde{y}) = 0$. This results in the corresponding $|\tilde{y}\rangle$ component in $|\Psi_3\rangle$ to have a full amplitude of 1:

$$\frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{f(x) \oplus x \cdot \tilde{y}} |\tilde{y}\rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} |\tilde{y}\rangle = |\tilde{y}\rangle. \quad (11.9)$$

It follows that $|\Psi_3\rangle$ solely contains $|\tilde{y}\rangle$; all other components nullify due to the normalization condition. Therefore, measuring $|\Psi_3\rangle$ invariably yields $|\tilde{y}\rangle$.

To comprehend why $f(x) = x \cdot \tilde{y}$ symbolizes a circular periodic function, consider $|\tilde{y}\rangle = |00\dots 01\dots\rangle$ with k initial 0s. The leading digits in $|x\rangle$ corresponding to the leading 0s in $|\tilde{y}\rangle$ do not influence $x \cdot \tilde{y}$, making the periodicity of $f(x)$ equal to 2^{n-k} .

4 Measurement in Alternative Basis

The measurements are not confined solely to computational basis states. Any set of 2^n orthonormal basis states is compatible with the generalized DJ algorithm. However, shifting to an alternative basis might yield functions lacking straightforward properties like apparent periodicity. Moreover, transitioning to an alternative basis necessitates an extra unitary transformation before measurement.

To illustrate, consider grouping the 2^n basis states into 2^{n-1} distinct pairs, which need not be consecutive. Given a pair, say $|i\rangle$ and $|j\rangle$, two orthogonal states can be derived: $\frac{1}{\sqrt{2}}(|i\rangle + |j\rangle)$ and $\frac{1}{\sqrt{2}}(|i\rangle - |j\rangle)$. Together, these states constitute a new orthogonal basis. Measuring against $\frac{1}{\sqrt{2}}(|i\rangle + |j\rangle)$ unequivocally identifies the function $f(x) = x \cdot (i \oplus j)$.

5 Spectral Measurement with the DJ Circuit

To further generalize the DJ algorithm, we leave the nature of the function $f(x)$ unconstrained. Employing repeated measurements on the DJ circuit, as shown in Fig. 11.1, allows us to collect data over numerous iterations. This process enables us to derive an empirical probability distribution for each of the 2^n possible outcomes and construct a length- 2^n vector representing the spectral distribution of the function $f(x)$.

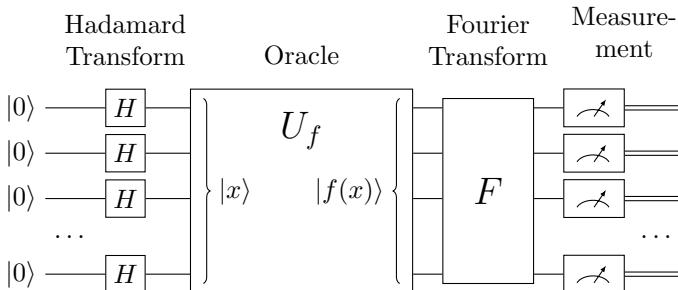


Figure 11.2: Quantum Circuit for Spectral Measurement

Furthermore, while the DJ algorithm traditionally employs the Hadamard transform to facilitate measurements of the spectral distribution in relation to square waves, a more sophisticated approach involves replacing the secondary Hadamard transform with the Fourier transform, denoted as F . This substitution provides a more accurate representation of the spectral density of $f(x)$, as visually depicted in Fig. 11.2. Moreover, we can extend the numerical range by substituting $f(x)$ with a function characterized by a wider output range, moving beyond the binary confines of $\{0, 1\}$.

The theoretical probability distribution, representing the spectral density of $f(x)$, is given by:

$$P_x(U_f) = |\langle x|FU_f|\Psi_u\rangle|^2, \quad (11.10)$$

where $|\Psi_u\rangle$ is the uniform superposition, as defined in Eq. 11.2, and $x \in \{0,1\}^n$, or equivalently, $x \in [0, 2^n - 1]$ in decimal representation.

DJ Algorithm Revisited

Now, let's analyze how Eq. 11.10 operates in the context of the DJ scenarios, where F is the Hadamard transform H_n .

- Constant functions

Based on Eq. 11.3, for the constant-0 function, $U_f|x\rangle = |x\rangle$ for any $|x\rangle$, leading to $U_f|\Psi_u\rangle = |\Psi_u\rangle$. Similarly, for the constant-1 function, we have $U_f|\Psi_u\rangle = -|\Psi_u\rangle$. Therefore, for constant functions, $U_f|\Psi_u\rangle = \pm|\Psi_u\rangle$. Since $H_n|\Psi_u\rangle = |0\rangle$, the equation simplifies to:

$$P_x(U_{\text{constant}}) = |\pm\langle x|H_n|\Psi_u\rangle|^2 = |\pm\langle x|0\rangle|^2 = \begin{cases} 1 & \text{for } x = 0, \\ 0 & \text{for } x \neq 0. \end{cases} \quad (11.11)$$

- Balanced functions

For a balanced function, $U_f|x\rangle = |x\rangle$ for half of the x values, and $U_f|x\rangle = -|x\rangle$ for the other half. Consequently, $U_f|\Psi_u\rangle$ becomes a vector, half of whose components are $\frac{1}{\sqrt{2^n}}$ and the other half $-\frac{1}{\sqrt{2^n}}$. We denote this state as $|\Psi_b\rangle$.

It's worth noting that $|\Psi_b\rangle$ has no overlap with $|\Psi_u\rangle$, which has all its components equal to $\frac{1}{\sqrt{2^n}}$. Furthermore, since $H_n|0\rangle = |\Psi_u\rangle$, we have:

$$P_x(U_{\text{balanced}}) = \begin{cases} |\langle\Psi_u|\Psi_b\rangle|^2 = 0 & \text{for } x = 0, \\ |\langle x|H_n|\Psi_b\rangle|^2 & \text{for } x \neq 0. \end{cases} \quad (11.12)$$

- Generalized DJ

In the generalized DJ, or equivalently, the BZ algorithm, where $f(x) = x \cdot \tilde{y}$ (Eq. 11.8), and $U_f|\Psi_u\rangle$ is represented by $|\Psi_3\rangle$ (as defined in Eq. 11.7), Eq. 11.10 can be rewritten as:

$$P_x(U_{\text{gen}}) = \left| \langle x| \frac{1}{2^n} \sum_{z,y \in \{0,1\}^n} (-1)^{(z \cdot \tilde{y}) \oplus (z \cdot x)} |y\rangle \right|^2 \quad (11.13a)$$

$$= \left| \frac{1}{2^n} \sum_{z \in \{0,1\}^n} (-1)^{(z \cdot \tilde{y}) \oplus (z \cdot x)} \right|^2 \quad (\because \langle x|y\rangle = 0 \text{ unless } y = x) \quad (11.13b)$$

$$= \begin{cases} 1 & \text{for } x = \tilde{y}, \\ 0 & \text{for } x \neq \tilde{y}. \end{cases} \quad (\because (z \cdot x) \oplus (z \cdot x) = 0) \quad (11.13c)$$

11.1.5 *The Phase Oracle

We now focus on the phase oracle, a quantum unitary realization of a specific instance of the binary function $f(x)$. This notion of a phase oracle extends to several quantum algorithms, not just DJ. Two critical facets require elucidation: the quantum realization of $f(x)$, and the phase-kickback mechanism.

1 Unitary Representation of a Function

Harnessing quantum parallelism necessitates that the oracle is constructed as a unitary transformation, capable of handling superposition states. We will exemplify this with a specific instance detailed in the succeeding exercise.

Exercise 11.6 It is known that the parity function $f(x) = \text{XOR}(x) = x_1 \oplus x_2 \oplus \dots \oplus x_n$ is a balanced function (see Problem 11.2).

Now establish that inverting any number of input bits in the above function results in another balanced function, for instance, $f(x) = \overline{x_1} \oplus x_2 \oplus \dots \oplus x_n$.

Figure 11.3 displays a circuit implementing the function $f(x) = \overline{x_1} \oplus x_2 \oplus x_3 \oplus \overline{x_4}$ across five qubits. The fifth qubit conveys the outcome, $|f(x)\rangle$. The operation \oplus is realized via CNOT gates, while X gates achieve the inversions.

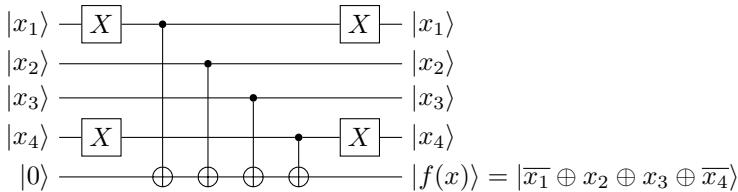


Figure 11.3: An Oracle for a Balanced Function

Exercise 11.7 Confirm that the circuit illustrated in Fig. 11.3 outputs $|f(x)\rangle$ on its auxiliary qubit, where $f(x) = \overline{x_1} \oplus x_2 \oplus x_3 \oplus \overline{x_4}$.

2 Phase Kickback

A limitation of the oracle in Fig. 11.3 is its dependence on an auxiliary qubit to relay the function result. For the DJ algorithm, the function result must be reflected as changes in the original qubits. This can be achieved by initializing the auxiliary qubit as $|-\rangle$, typically done with an X gate followed by an H gate, as depicted in Fig. 11.4.

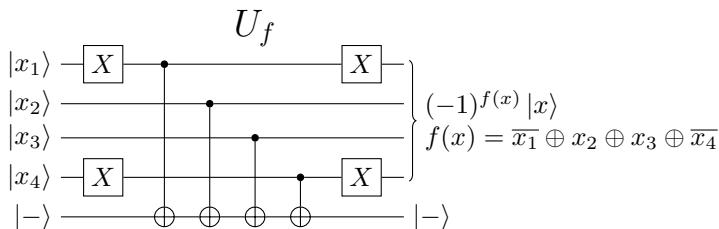


Figure 11.4: A Phase Oracle for a Balanced Function

This adjustment ensures the auxiliary qubit remains $|-\rangle$, but introduces a phase of -1 to the original qubits if $f(x) = 1$, a phenomenon termed as phase kickback.

The root cause behind this behavior stems from the properties of the CNOT gate (see § 7.1.1):

$$\text{CNOT } |0\rangle |-\rangle = |0\rangle |-\rangle, \quad (11.14\text{a})$$

$$\text{CNOT } |1\rangle |-\rangle = -|1\rangle |-\rangle. \quad (11.14\text{b})$$

Consequently, the state of any input, $|x_i\rangle$, transitions to $(-1)^{x_i} |x_i\rangle$.

Following all the CNOT operations, the entire system reaches the state:

$$(-1)^{\overline{x_1}} |\overline{x_1}\rangle (-1)^{x_2} |x_2\rangle (-1)^{x_3} |x_3\rangle (-1)^{\overline{x_4}} |\overline{x_4}\rangle |-\rangle.$$

Subsequent X gates negate the inversion effects on $\overline{x_1}$ and $\overline{x_4}$. This results in:

$$(-1)^{f(x)} |x\rangle |-\rangle.$$

The phase kickback is now complete.



Though the phase $(-1)^{f(x)}$ may seemingly be aligned with either $|x\rangle$ or $|-\rangle$, its dependence on x restricts the alignment with the former for superposition states. For instance:

$$U_f \frac{1}{\sqrt{2}} (|0000\rangle + |1000\rangle) |-\rangle = \frac{1}{\sqrt{2}} (|0000\rangle - |1000\rangle) |-\rangle.$$



Given its limited utility within the oracle, the auxiliary qubit, commonly termed an ancilla qubit, can be disregarded outside the oracle.

The phase oracle can also be implemented with CZ gates instead of CNOTs, as showcased in Fig. 11.5.

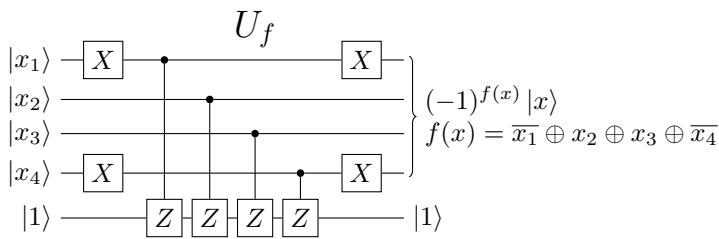


Figure 11.5: A Phase Oracle Using CZ Gates

11.2 QUBO, VQE, QAOA, and AQC

In an era where computational demands continually escalate and classical computing paradigms show limitations, the exploration of novel computational methods has become imperative. As we transition from limited Noisy Intermediate-Scale Quantum (NISQ) devices to the burgeoning era of quantum utility, these quantum systems

emerge as a promising avenue. Despite their limitations, NISQ and emerging utility-era quantum computers exhibit potential to outperform classical counterparts in areas like optimization, quantum simulation, and machine learning[23].

Quantum Computing Applications in the NISQ/Unity Era

In the NISQ and emerging quantum utility era, quantum computing, particularly through hybrid quantum-classical algorithms, is poised to benefit real-world applications in several key areas:

- Optimization: Essential for complex problem-solving in finance, logistics, and more.
- Quantum Simulation: Useful in material science, drug discovery, and battery efficiency.
- Machine Learning: Enhancing data processing, pattern recognition, and predictive analytics.

Due to the limited scope of this text, our focus will be on Optimization. Quantum simulation requires an advanced understanding of quantum mechanics, and quantum machine learning necessitates a comprehensive knowledge of classical machine learning.

Central to this pursuit is the Quadratic Unconstrained Binary Optimization (QUBO) problem. As a combinatorial optimization paradigm, QUBO finds applications spanning a broad spectrum from finance and economics to advanced machine learning endeavors. Owing to its NP-hard nature, several classical challenges from theoretical computer science, such as max-cut, graph coloring, and the partition problem, have been mapped into the QUBO framework. Furthermore, this mapping extends to sophisticated machine learning models like support-vector machines, clustering algorithms, and probabilistic graphical models.

NP-hard Problems

Loosely speaking, the term NP-hard designates a class of problems that are particularly challenging to solve in polynomial time. This means that as the size of the problem increases, the time required to find an exact solution grows much more rapidly than a polynomial function of the problem size. These types of problems frequently arise in various real-world applications, such as logistics and cryptography, where finding an optimal solution is crucial yet computationally demanding.

In the quest to solve QUBO, two prominent quantum-classical hybrid algorithms emerge: the Quantum Approximate Optimization Algorithm (QAOA) and the Variational Quantum Eigensolver (VQE). The QAOA leverages quantum mechanics to generate approximate solutions to combinatorial problems by alternating between unitary transformations. On the other hand, VQE aims to find the ground state energy of a given Hamiltonian by using a parameterized quantum circuit and iteratively refining its parameters.

Moreover, given QUBO's intrinsic relationship with Ising models, it is recognized as a key problem class in adiabatic quantum computation (AQC). Herein, solutions are procured through a specialized physical process known as quantum annealing.

To elucidate the core principles and subtleties of these algorithms and com-

putational strategies, we will hone in on the Max-Cut problem as our illustrative example.

Achieving Quantum Advantages with VQE and QAOA

VQE (Variational Quantum Eigensolver) and QAOA (Quantum Approximate Optimization Algorithm) exhibit three key traits that make them promising for realizing quantum advantages, particularly on near-term, noisy, intermediate-scale quantum (NISQ) devices: hybrid nature, efficiency, and adaptability.

The hybrid structure of VQE and QAOA enables effective utilization of both classical and quantum resources, making them well-suited for NISQ devices. These algorithms partition the computational workload between a classical optimizer and a quantum processor. This division allows for a fallback to classical techniques for error mitigation and tasks that may not yet be efficiently manageable on current quantum hardware. Such a distributive strategy is particularly beneficial during the transitional phase leading to fully fault-tolerant quantum computing.

Efficiency in VQE and QAOA manifests through their judicious use of quantum resources and their capability to yield good approximations using limited qubits and shallow circuits. Both algorithms employ parameterized quantum circuits with depths usually falling within the coherence times of existing NISQ systems. This minimizes the error probabilities associated with longer computations and maximizes the computational capabilities of current quantum hardware, an essential feature for near-term applications.

The adaptability of these algorithms stems from their iterative mechanisms and parameter fine-tuning capabilities. This not only confers robustness against quantum errors but also offers flexibility in addressing a wide range of optimization and eigenvalue problems. Both VQE and QAOA can be adapted to specific problem characteristics, enabling the inclusion of specialized heuristics or alternative cost functions. This level of adaptability enhances their applicability across various domains, further substantiating their relevance in the context of near-term quantum computing.

A variety of concepts and acronyms are introduced in this section. For easy reference, see Table 11.2.

11.2.1 The Max-Cut Problem

The Max-Cut problem, originated from graph theory, is an exemplary QUBO problem. Given an undirected graph with a set of vertices and edges, the objective is to partition the vertices into two disjoint sets (or equivalently, color the vertices with two distinct colors) such that the number of edges connecting these sets (or having differently colored endpoints) is maximized. This scenario is illustrated in Fig. 11.6 comprising five vertices.

A useful analogy is visualizing each vertex as a city and each edge as a road connecting two cities. The Max-Cut problem can then be understood as optimally segregating these cities into two regions to maximize the number of interconnecting roads.

Another analogous representation is the influence maximization challenge in social networks. Envision a system of interconnected individuals influencing one another. Each individual corresponds to a vertex, while their interactions are represented by

Term	Description	Relationships
QUBO	Quadratic Unconstrained Binary Optimization. Problems aiming to minimize or maximize an objective function that comprises quadratic terms in binary variables.	Can be represented by the Ising model. Max-cut and TSP are examples.
Max-Cut	Optimization problem in graph theory.	Can be formulated as a QUBO or Ising Model.
TSP	Traveling Salesman Problem	Can be formulated as a QUBO or Ising Model.
VQA	Variational Quantum Algorithms. A class of algorithms that leverage variational principles combined with ansatzes.	VQE and QAOA are examples.
VQE	Variational Quantum Eigensolver. An algorithm for finding the ground state of a Hamiltonian.	A special case of VQA.
QAOA	Quantum Approximate Optimization Algorithm. Inspired by adiabatic quantum evolution.	A special case of VQA.
Ansatz	Parameterized trial state vector implemented as a quantum circuit.	Component of VQE and QAOA.
Ising Model	Model describing interactions of $\frac{1}{2}$ -spins in magnetic fields.	Can represent QUBO problems, including Max-Cut and TSP.
AQC	Adiabatic Quantum Computing. Computation model based on initializing a system in the ground state of a simple Hamiltonian and then slowly evolving the Hamiltonian to one representing the problem.	Includes Quantum Annealing. Inspired QAOA.
Quantum Annealing	Optimization technique using quantum fluctuations to escape local minima. Implemented on dedicated hardware (quantum annealers).	Subset of AQC.

Table 11.2: Relationships Between Terms Related to Quantum Optimization

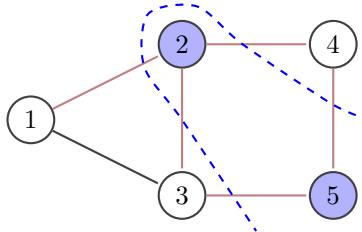


Figure 11.6: Max-Cut Problem Example

edges. This model effectively captures marketing dynamics where the strength of influence between individuals, represented by edge weights, governs their purchasing decisions. For a marketing strategy offering free products to select individuals, the central goal becomes determining the ideal subset of these individuals to optimize revenue.

Being NP-hard, the Max-Cut problem implies that as graph size grows, the computational time to determine the precise solution can expand exponentially, rendering classical methods ineffective for substantial graphs. Nonetheless, due to its significance as a representative for numerous optimization challenges, this problem, despite its deceptive simplicity, underscores the computational intricacies inherent in obtaining optimal solutions, establishing it as a prime candidate for quantum computational techniques such as QAOA and quantum annealing.

1 Problem Formulation

The weighted max-cut problem can be formally defined as follows:

For an undirected graph $G = (V, E)$ comprising n nodes, where V denotes the vertex set and E symbolizes the edge set, edge weights are given by w_{ij} for $(i, j) \in E$. A cut bifurcates the primary set V into two subsets, labelled 0 and 1, respectively. The function to optimize, in this context, is the aggregate weights on edges connecting distinct subsets across the cut. By attributing $x_i = 0$ or $x_i = 1$ to each node i , the objective becomes maximizing the global profit function:

$$\tilde{C}(x) = \sum_{(i,j) \in E} w_{ij}(x_i \oplus x_j), \quad (11.15)$$

where \oplus denotes the XOR function. Specifically, $x_i \oplus x_j = 1$ if $x_i \neq x_j$, meaning that nodes i and j belong to different subsets; the value is 0 otherwise.

In the simplified marketing archetype, w_{ij} characterizes the likelihood that individual j will purchase a product if i has received it as a free sample. Notably, the value of w_{ij} can exceed 1 or even be negative. A value greater than 1 could reflect scenarios where individual j is motivated to buy multiple products, while a negative value could indicate dissuasion. The objective is to maximize the cumulative purchasing probabilities, which equates to revenue optimization. If the anticipated profit probability exceeds the initial sample costs, the strategy is considered profitable.

In an extended model, nodes themselves possess weights, denoted by w_i , mirroring the probability that an individual receiving a free sample repurchases the product. Incorporating this facet, the objective function evolves to:

$$\tilde{C}(x) = \sum_{(i,j) \in E} w_{ij}(x_i \oplus x_j) + \sum_{i \in V} w_i x_i. \quad (11.16)$$

i

The objective function for the extended Max-Cut problem (Eq. 11.16) is mathematically equivalent to that of a Quadratic Unconstrained Binary Optimization (QUBO) problem.

Exercise 11.8 Formulate the profit function $\tilde{C}(x)$ for the max-cut problem in Fig. 11.6, assuming $w_{ij} = 1$ and $w_i = 0$.

11.2.2 *The Ising Model

The Ising model, introduced in § 6.5.3, describes a system of spins under the influence of a transverse magnetic field. It provides a bridge between physical systems and computational problems, allowing the use of quantum devices to solve problems of practical interest.

To derive a solution to the weighted max-cut problem via a quantum computer, an initial step is the mapping of this problem to an Ising Hamiltonian. We map the binary variables $x_i, x_j \in \{0, 1\}$ in Eq. 11.16 to $z_i, z_j \in \{-1, 1\}$ (since the eigenvalues of Z are $\{-1, 1\}$):

$$x_i \rightarrow \frac{1}{2}(1 - z_i), \quad x_i \oplus x_j \rightarrow \frac{1}{2}(1 - z_i z_j). \quad (11.17)$$

Replacing z_i and z_j with the corresponding Pauli operators Z_i and Z_j , we deduce that:

$$C(Z) = \sum_{(i,j) \in E} \frac{w_{ij}}{2} (I - Z_i Z_j) + \sum_{i \in V} \frac{w_i}{2} (I - Z_i) \quad (11.18a)$$

$$= -\frac{1}{2} \sum_{(i,j) \in E} w_{ij} Z_i Z_j - \frac{1}{2} \sum_{i \in V} w_i Z_i + \text{const}, \quad (11.18b)$$

where const is a Z -independent term and can be ignored.

i

$C(Z)$ is a matrix while $\tilde{C}(x)$ is a number.

Hence, the weighted max-cut problem can be equated to minimizing the Ising Hamiltonian:

$$H = \frac{1}{2} \sum_{(i,j) \in E} w_{ij} Z_i Z_j + \frac{1}{2} \sum_{i \in V} w_i Z_i, \quad (11.19)$$

which corresponds to identifying the ground state of the spin system portrayed by the Ising model.



In the above notation, $Z_i Z_j$ is a shorthand for $Z_i \otimes Z_j$, acting on qubits i and j .

The eigenvalues $+1$ and -1 of the Z operator signify the two separate sets into which the vertices are partitioned. When the $Z_i Z_j$ product equals -1 , this denotes that vertices i and j are placed in different sets, and the corresponding edge is a *cut edge*. Conversely, if $Z_i Z_j$ results in $+1$, this implies that vertices i and j belong to the same set, and the edge is an *uncut edge*.

Exercise 11.9 Express Eq. 11.19 explicitly as an 8×8 matrix, for a system of three nodes on a triangle, assuming $w_{ij} = 1$ for all edges and $w_i = 1$ for all nodes.

Exercise 11.10 Formulate the Hamiltonian H for the max-cut problem in Fig. 11.6, assuming $w_{ij} = 1$ and $w_i = 0$.

The Ising model can subsequently be tackled using a diverse array of quantum algorithms, including the Quantum Approximate Optimization Algorithm (QAOA) and Variational Quantum Eigensolver (VQE), or via quantum annealing apparatuses like the D-Wave systems.

Applications of the Ising Model

The applications of the Ising model are extensive, surpassing just the Max-Cut:

Combinatorial Optimization Problems: Beyond the max-cut problem explored here, challenges such as the traveling salesman problem and other NP-hard problems can be transposed onto an Ising framework.

Machine Learning: The Ising model's reach extends to quantum machine learning, particularly in modeling Boltzmann machines, which are identified as stochastic recurrent neural networks. Herein, the parameters of the Ising model encapsulate the weights and biases of the Boltzmann machine, and the model's equilibrium state offers a probabilistic representation of the data.

Quantum Error Correction: Grounded in the Ising model, topological quantum error correction codes emerge as promising candidates for trustworthy quantum computation. The toric code, a two-dimensional Ising variant with additional constraints, facilitates the comprehension and realization of quantum error correction.

Condensed Matter Physics: The quantum version of the Ising model plays an instrumental role in simulating quantum phase transitions, a subject capturing significant interest within the domain of condensed matter physics. More specifically, the Ising model provides insights into the transitional phases between distinct states of matter induced by quantum perturbations. This encompasses the exploration of phenomena like ferromagnetism, anti-ferromagnetism, and other unconventional states of matter.

11.2.3 *Variational Quantum Eigensolver (VQE)

The Variational Quantum Eigensolver (VQE) is a hybrid quantum-classical algorithm designed to harness the capabilities of quantum computers. While its initial development targeted quantum chemistry applications—often considered the most practical use-case for Noisy Intermediate-Scale Quantum (NISQ) computers—the algorithm's

scope extends beyond this, making it also suitable for solving optimization problems.

At the heart of VQE is the mapping of the problem onto a Hamiltonian. The ground state of this Hamiltonian encodes the solution. The variational principle ensures that for any chosen trial state, $\psi(\theta)$, its energy expectation value will always be greater than or equal to the ground state energy, E_0 . Thus, by iteratively optimizing the parameters of the trial state, the VQE approximates the ground state energy and the associated quantum state.

1 The Variational Principle

The variational principle is a fundamental concept in quantum mechanics, formally stated as:

$$\langle \psi(\theta) | H | \psi(\theta) \rangle \geq E_0, \quad (11.20)$$

where E_0 represents the ground state energy.

This principle lays the groundwork for the VQE algorithm. In quantum systems with many qubits, the state space becomes intractably large for exhaustive energy minimization. However, a quantum computer can efficiently prepare trial states and evaluate their corresponding energy expectation values. Following these quantum computations, a classical optimizer adjusts the parameters θ in order to minimize the energy, thus approximating the ground state of the system.



When maximizing a profit function is the objective, it is equivalent to minimizing the negative of that function.

2 Ansatz

The ansatz in VQE refers to the trial quantum state, $\psi(\theta)$, parameterized by an array of parameters θ . It is an educated guess of the solution, based on certain heuristics or insights. Implemented as a parameterized quantum circuit, the quality of the ansatz is crucial to the accuracy of the VQE. If the chosen ansatz is not expressive enough to approximate the ground state of the problem, the optimization is unlikely to yield accurate results.

3 Measurement

The ansatz is run repeatedly for m times. For each run, the n qubits are measured in the computational basis; the k -run yields n measurement values $z_1^{(k)}, z_2^{(k)}, \dots, z_n^{(k)}$, where $z_i^{(k)} = 1$ for measurement outcome $|0\rangle$, and -1 for $|1\rangle$. Essentially, we are measuring each qubit with the Z observable.

Then the expected value of the Hamiltonian is estimated according to Eq. 11.19:

$$\langle \psi | H | \psi \rangle = \frac{1}{2m} \sum_{k=1}^m \sum_{i \in V} w_i z_i^{(k)} + \frac{1}{2m} \sum_{k=1}^m \sum_{(i,j) \in E} w_{ij} z_i^{(k)} z_j^{(k)}. \quad (11.21)$$

Note the above calculation is only valid for the Ising model, where the Hamiltonian involves Z solely and is hence diagonal in the computational basis. The general case is more complicated, to be discussed at the end of this subsection.

4 Iteration Procedure

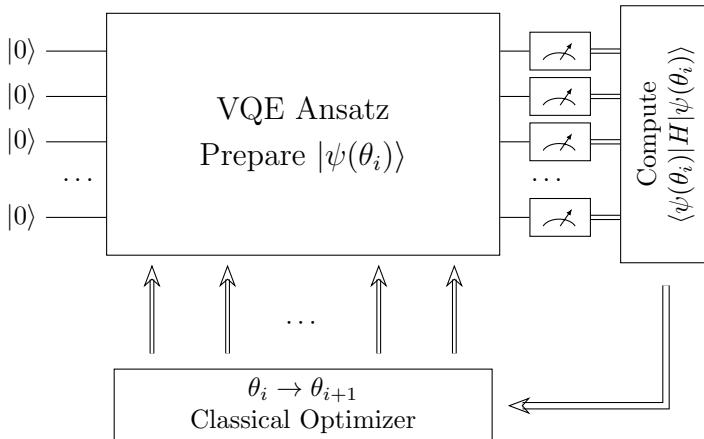


Figure 11.7: VQE Ansatz and Iteration Procedure

The VQE follows an iterative procedure (Fig. 11.7) comprising:

1. Crafting a parameterized quantum circuit that serves as the ansatz, $|\psi(\theta)\rangle$.
2. Measuring the ansatz qubits on a quantum computer and processing the outcomes classically to estimate the expected value $\langle\psi(\theta)|H|\psi(\theta)\rangle$.
3. Employing a classical optimizer to propose adjustments to the ansatz's parameters.

The iterations continue until convergence, or when the change in the expected value is below a predefined threshold.



The weights $w_{i,j}$ are not directly incorporated into the variational parameters θ ; instead, θ is utilized to construct the trial ground state vector within the quantum circuit framework. The weights $w_{i,j}$ are used to compute the expected value of the Hamiltonian on a classical computer. An optimization routine, executed on a classical computer, then uses this calculated expected value to iteratively adjust θ . This process aims to minimize the expected value of the Hamiltonian, thereby accurately approximating the ground state with an optimal θ .

5 *General Case

In the general case, such as when VQE is used in chemistry simulation, computation of the expected value $\langle\psi(\theta)|H|\psi(\theta)\rangle$ is non-trivial. The Hamiltonian must first be decomposed into Pauli strings, P_i :

$$H = \sum_i c_i P_i, \quad P_i \in \{I, X, Y, Z\}^{\otimes n}. \quad (11.22)$$

With the Pauli basis being orthogonal and having a norm of 2, the coefficients c_i can be expressed as:

$$c_i = \frac{1}{2^n} \text{tr}(HP_i). \quad (11.23)$$

Consequently, the expected value of H is:

$$\langle \psi | H | \psi \rangle = \sum_i c_i \langle \psi | P_i | \psi \rangle. \quad (11.24)$$

Sampling the Pauli Operators

When dealing with a Hamiltonian H that comprises not just the Z operator but also X and Y , distinct measurement strategies are employed. Direct measurements in the computational basis suffice for evaluating $\langle \psi | Z | \psi \rangle$. However, to obtain expected values for X and Y , one must rotate the basis prior to measurement, using the Hadamard gate for X and $R_x(-\frac{\pi}{2})$ for Y , as described in § 3.4.5.

In this more general context, calculating the overall expected value $\langle \psi | H | \psi \rangle$ involves measuring each term in the Hamiltonian's Pauli basis expansion (Eq. 11.24). Each term generally necessitates a unique run of the quantum circuit (or ansatz), meaning up to 4^n runs could be required in the worst-case scenario. However, many practical quantum problems feature Hamiltonians with far sparser representations, thereby reducing the number of circuit runs needed to achieve a reliable estimation of the Hamiltonian's expected value.

6 The Ground State

For the Max-Cut problem, since the Hamiltonian is diagonal in the computational basis, its eigenstates will also be computational basis states. Hence, the ground state that minimizes the cost function will correspond to a computational basis state.

In the general case, this may no longer be the case. The ground state may be a superposition state in the computational basis.

11.2.4 *Quantum Approximate Optimization Algorithm (QAOA)

The Quantum Approximate Optimization Algorithm (QAOA) is based on the Variational Quantum Algorithm (VQA) framework. It is specifically designed to solve combinatorial optimization problems such as Max-Cut. The primary objective is to identify the quantum state that minimizes the expected value of a Hamiltonian which encapsulates the cost function of the problem at hand.

While QAOA shares the iterative nature of the Variational Quantum Eigensolver (VQE), it employs a specialized ansatz, as depicted in Fig. 11.8.

1 Structure of the QAOA Ansatz

The QAOA ansatz employs an alternating sequence of unitary transformations, which are derived from two Hamiltonians: the cost Hamiltonian H_C and a mixing Hamiltonian H_B . Specifically, the ansatz consists of repeated blocks of two types of unitary transformations, $U_C(\gamma)$ and $U_B(\beta)$, expressed as:

$$U_B(\beta) = e^{-i\beta H_B}, \quad U_C(\gamma) = e^{-i\gamma H_C}, \quad (11.25)$$

where β and γ are variational parameters optimized through classical means.

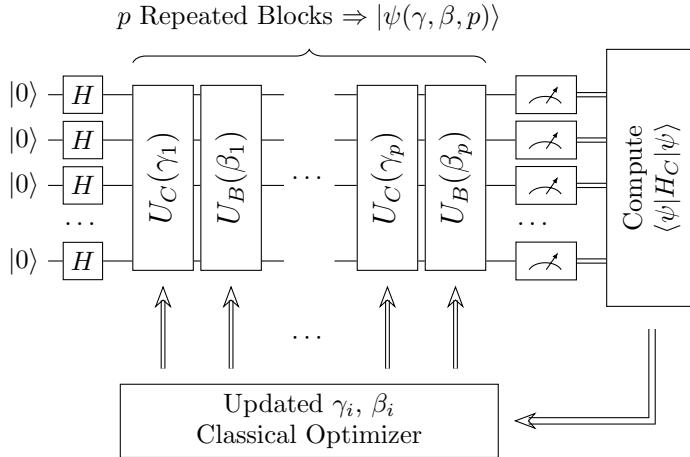


Figure 11.8: QAOA Ansatz and Iterative Procedure

2 Cost Hamiltonian

The cost Hamiltonian, H_C , encodes the objective function of the optimization problem. It assigns an “energy” or “cost” to specific configurations of the system. The goal for combinatorial optimization problems is to find the configuration that minimizes this energy. A representative expression for H_C in the context of Max-Cut is:

$$H_C = \frac{1}{2} \sum_{(i,j) \in E} w_{ij} Z_i Z_j + \frac{1}{2} \sum_{i \in V} w_i Z_i, \quad (11.26)$$

Here, V and E are the sets of vertices and edges in the graph $G = (V, E)$. w_{ij} and w_i are the weights for edges and vertices, and Z_i is the Pauli-Z operator on the i -th qubit.

3 Mixing Hamiltonian

The mixing Hamiltonian H_B facilitates transitions between different cut configurations, thereby aiding the system in exploring the space of potential solutions. A common choice for H_B is:

$$H_B = \sum_i X_i, \quad (11.27)$$

where X_i is the Pauli-X operator on the i -th qubit.

4 Interplay and Iteration

The interplay between H_C and H_B lies at the heart of QAOA’s functionality. While H_C encodes problem-specific penalties or rewards, H_B keeps the quantum state in a superposition of various configurations. By iteratively applying these Hamiltonians and fine-tuning the variational parameters, QAOA aims to direct the quantum system toward H_C ’s ground state. The optimal depth p is generally determined

through empirical studies, balancing the need for algorithmic performance against computational resources. This ground state corresponds to the problem's optimal or near-optimal solution.

5 Example Application to Max-Cut

Consider a simple Max-Cut problem depicted in Fig. 11.6. The graph has vertices $V = \{1, 2, 3, 4, 5\}$ and edges $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 5), (4, 5)\}$, with all edge weights set to unity, and node weights to zero.

The solution is obviously a cut shown by the dashed curve: with only one un-cut edge $(1, 3)$, and nodes partitioned into $\{1, 3, 4\}$ and $\{2, 5\}$. But we will walk through the QAOA process as a demonstration of its concepts and procedure.

Its cost Hamiltonian can be expressed using the Ising model as:

$$H_C = \frac{1}{2} \sum_{(i,j) \in E} Z_i Z_j. \quad (11.28)$$

When the product of $Z_i Z_j$ is -1 , it means the vertices i and j are in different sets (a cut edge), contributing negatively to the total energy. Therefore, minimizing $\langle H_C \rangle$ corresponds to maximizing the number of cuts, the goal of the Max-Cut problem.

The unitary block $U_C(\gamma)$ is given by:

$$U_C(\gamma) = e^{-i\gamma H_C} = \prod_{(i,j) \in E} e^{-i\frac{\gamma}{2} Z_i Z_j} = \prod_{(i,j) \in E} ZZ_{ij}(\frac{\gamma}{2}), \quad (11.29)$$

where each factor in the product is a $ZZ(\frac{\gamma}{2})$ gate (see § 7.1.5) acting on qubits i and j .

The unitary block $U_B(\beta)$ is given by:

$$U_B(\beta) = e^{-i\beta H_B} = \prod_{i \in V} e^{-i\beta X_i} = \prod_{i \in V} R_{x_i}(\beta), \quad (11.30)$$

where each factor in the product is an $R_x(\beta)$ gate acting on qubit i .

The building block for the QAOA ansatz comprising $U_C(\gamma)$ and $U_B(\beta)$ is shown in Fig. 11.9.

After sufficient number of iterations, the algorithm arrives at a ground state expected value $\langle H_C \rangle = -2$. The resulting ground state is $|01001\rangle$ (or equivalently $|10110\rangle$), yielding a partition of $\{1, 3, 4\}$ and $\{2, 5\}$ as anticipated.

Exercise 11.11 Formulate the cost function and the Ising Hamiltonian for the six-node Max-Cut problem shown in the figure below. Assume all the edge weights are 2. Node weights are 1 for even nodes and -1 for odd nodes. Solve the problem to find the optimal cut using QAOA, implemented in Qiskit, Cirq or a similar platform. Provide the final (i.e., most probable) state vector and the measure of the quality of the cut.

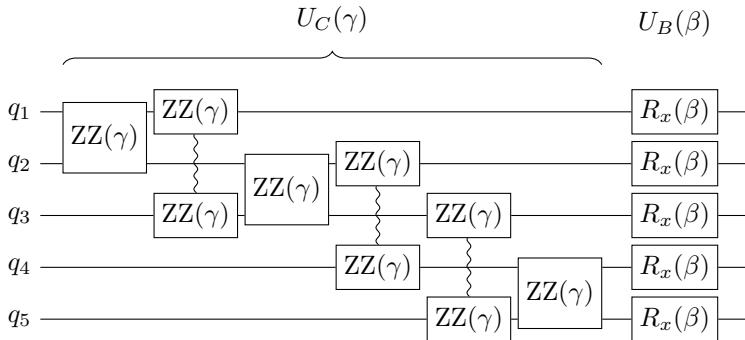
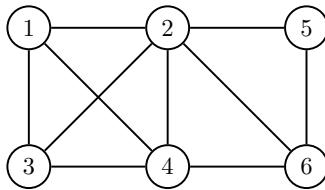


Figure 11.9: Example of a QAOA Ansatz Building Block



11.2.5 *Traveling Salesman Problem

The Traveling Salesman Problem (TSP) is a seminal problem in the realm of QUBO. It involves finding the shortest route through a list of cities, ensuring that each city is visited exactly once before returning to the starting city. This can be represented as a weighted graph where the cities are vertices and the distances between them serve as weighted edges, as shown in Fig. 11.10. Much like the Max-Cut problem, TSP is NP-hard, meaning that computational time grows substantially with an increase in the number of cities.

Owing to its wide range of applications in fields such as logistics and transportation, the TSP has attracted attention from both classical and quantum optimization techniques. Despite its seeming simplicity, the problem is deeply rooted in combinatorial optimization, underlining the value of quantum approaches for solving complex problems.

1 Problem Formulation

The TSP is mathematically framed as follows: one seeks to minimize the tour distance in a graph $G = (V, E)$, where n nodes represent cities and d_{ij} denotes the distance between nodes i and j . The tour distance for a cycle is defined as:

$$\tilde{L}(x) = \sum_{i=1}^n \sum_{j=1, j \neq i}^n d_{ij} x_{ij}, \quad (11.31)$$

where $x_{ij} \in \{0, 1\}$ indicates if the tour goes from node i to j , with $x_{ij} = 1$ if so.

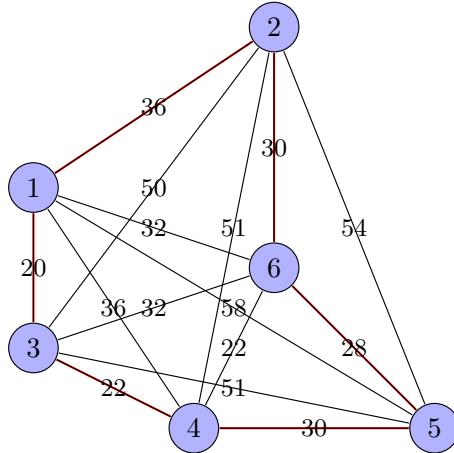


Figure 11.10: Example of the Traveling Salesman Problem

2 Constraints

To ensure that each city is visited exactly once, the following constraints are imposed:

$$S_j \equiv \sum_{i,i \neq j} x_{ij} = 1 \quad \text{for } j = 1, 2, \dots, n, \quad (11.32a)$$

$$S_i \equiv \sum_{j,j \neq i} x_{ij} = 1 \quad \text{for } i = 1, 2, \dots, n. \quad (11.32b)$$

The first equation states that for each city j , there must be exactly one city i such that $x_{ij} = 1$. This ensures that a path enters each city exactly once.

The second equation states that for each city i , there must be exactly one city j such that $x_{ij} = 1$. This ensures that a path exits each city exactly once.

3 Objective Function

The objective function, which includes both the tour distance and the constraints, is:

$$\tilde{C}(x) = \tilde{L}(x) + A \sum_{j=1}^n (1 - S_j)^2 + A \sum_{i=1}^n (1 - S_i)^2, \quad (11.33)$$

with A as a free parameter large enough to enforce the constraints, generally $A > \max(d_{ij})$.

This combined objective function is quadratic in x_{ij} :

$$\tilde{C}(x) = \sum_{i,j,k,l} w_{ijkl} x_{ij} x_{kl} + \sum_{ij} w_{ij} x_{ij}. \quad (11.34)$$

4 Quantum-Classical Hybrid Solution

The TSP can also be solved using hybrid quantum-classical algorithms like the Quantum Approximate Optimization Algorithm (QAOA). Due to the problem's

complexity, the above TSP implementation requires n^2 qubits, as opposed to the n qubits needed for Max-Cut. Modern quantum computing platforms often provide libraries to simplify this formulation.

For a more in-depth understanding, the reader is directed to the following references:

- Efficient traveling salesman problem solvers using the Ising model with simulated bifurcation [98].
- Clustering approach for solving traveling salesman problems via Ising model-based solver [37].

Exercise 11.12 Work through the six-node TSP example shown in Fig. 11.10. Formulate the cost function and the Ising Hamiltonian, and then solve the problem using QAOA. Use Qiskit, Cirq or a similar platform for implementation. Provide the final (i.e., most probable) state vector and the shortest tour distance. (If your platform does not have the capacity for simulating TSP with six nodes, reduce the number to five or four.)

11.2.6 Adiabatic Quantum Computation and Annealing

Quantum annealing is a specialized approach tailored for optimization problems. We will discuss this further using Max-Cut as an example.

1 From Universal Quantum Computing to Quantum Annealing

In the ever-evolving landscape of quantum computing, two primary paradigms have gained prominence: gate-based universal quantum computing and quantum annealing.

Gate-based universal quantum computers employ a series of quantum gates to control qubits. This model's universality makes it apt for diverse computational problems, from integer factorization to quantum system simulations. Algorithms like Deutsch-Jozsa, VQE, and QAOA, as well as Shor's and Grover's algorithms, belong to this paradigm. Some algorithms are not discussed here due to the scope of this text.

Adiabatic quantum computation (AQC) is another form of quantum computing, as introduced in § 4.4. Here, the problem is mapped to a quantum system whose ground state represents the solution. Starting from an initial, easily-understandable Hamiltonian, the system is slowly evolved to a final Hamiltonian that encodes the solution.

Quantum annealing is a subset of AQC tailored for optimization problems. It avoids the need for precise quantum gate sequences, allowing the system to evolve under its natural quantum mechanics.

This natural evolution makes quantum annealing particularly useful for optimization problems in various domains like machine learning, finance, and logistics. The lack of need for exact gate sequences can ease the construction of larger quantum systems, although some problems may not be well-suited for this approach.

2 Solving Max-Cut with Quantum Annealing

In AQC, we begin with the quantum system in its ground state, corresponding to the lowest energy eigenstate of an initial Hamiltonian H_{initial} . The Hamiltonian is then evolved towards a target H_{target} , which encodes the optimization problem. The evolving Hamiltonian, $H(t)$, is given by:

$$H(t) = (1 - s(t))H_{\text{initial}} + s(t)H_{\text{target}}, \quad (11.35)$$

where $s(t)$ transitions from 0 to 1 during the annealing process. According to the adiabatic theorem, a slow evolution ensures the system remains in its ground state, facilitating a smooth transition from H_{initial} to H_{target} and ultimately arriving at the solution.

A commonly used H_{initial} in quantum annealing is the transverse-field Ising Hamiltonian, given by:

$$H_{\text{initial}} = -\Gamma \sum_{i=1}^n X_i, \quad (11.36)$$

where Γ represents the strength of the transverse magnetic field, and X_i is the Pauli X operator acting on the i -th qubit.

The corresponding ground state is a uniform superposition of all computational basis states, represented as:

$$|\psi_{\text{initial}}\rangle = \bigotimes_{i=1}^n |+\rangle = \sum_{i=0}^{n-1} |i\rangle. \quad (11.37)$$

The target Hamiltonian, H_{target} , is given by:

$$H_{\text{target}} = \frac{1}{2} \sum_{(i,j) \in E} w_{ij} Z_i Z_j + \frac{1}{2} \sum_{i \in V} w_i Z_i, \quad (11.38)$$

where $Z_i Z_j$ represents $Z_i \otimes Z_j$, with Pauli Z acting on qubits i and j . (See Eq. 11.19.)

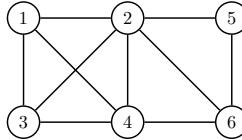
Once these Hamiltonians are defined, the quantum annealing machine takes over the computation. The annealing process initiates by preparing the quantum system in the ground state of H_{initial} . The system then evolves under the influence of the time-dependent Hamiltonian $H(t)$, which is a mixture of the initial and target Hamiltonians controlled by the parameter $s(t)$.

The annealing schedule, often predefined, controls the evolution of $s(t)$ from 0 to 1. As the system evolves, it naturally explores the energy landscape, guided by quantum fluctuations. The slow change ensures that the system remains near its ground state throughout the process, due to the adiabatic theorem.

At the end of the annealing schedule, the quantum state of the system approximates the ground state of H_{target} . This state is then measured, yielding a classical bit string that represents a near-optimal solution to the encoded optimization problem. Subsequent repetitions of this process can improve the quality of the solution through statistical sampling.

It's worth mentioning that while the annealing process is heuristic in nature, it has shown promising results in various optimization problems, offering a complementary approach to gate-based quantum computing.

Exercise 11.13 Solve the six-node Max-Cut problem shown in the figure below using a quantum annealing platform such as [D-Wave NetworkX](#). Assume all the edge weights are 2. Node weights are 1 for even nodes and -1 for odd nodes. Provide the final state vector and calculate the value of the cut as the measure of its quality. Compare your solution with the QAOA counterpart.



Exercise 11.14 Solve the six-node traveling salesman problem shown in Fig. 11.10 using a quantum annealing platform such as [D-Wave NetworkX](#). Provide the final state vector and the shortest tour distance. Compare your solution with the QAOA counterpart. (If your platform does not have the capacity for simulating TSP with six nodes, reduce the number to five or four.)

3 Conditions for Quantum Speedup and Limitations

In adiabatic quantum computation, and more specifically in quantum annealing, as H_{initial} evolves towards H_{target} (see Eq. 11.35), the energy gap between the ground state and the first excited state of the quantum system varies over time. This gap can become very small, potentially even reaching zero, a point known as a level crossing. The efficiency of the algorithm is closely tied to the behavior of the minimum energy gap, denoted as G .

The required annealing time τ is inversely proportional to the square of G , expressed as $\tau \propto 1/G^2$. Should G decrease exponentially with the size of the problem, or the number of qubits n , in the form $G \propto e^{-\alpha n}$ for some positive constant α , then τ will increase exponentially. This exponential increase in time negates any potential quantum speedup, as the system would then require an impractically long time to evolve.

For quantum annealing to be efficient and to offer a potential quantum speedup, the energy gap G must decrease polynomially with the problem size n . If G scales as $G \propto 1/n^k$ for some constant k , the annealing time τ would then increase only polynomially, remaining manageable and enabling a substantial quantum speedup compared to classical algorithms.

It is important to recognize that the behavior of the energy gap G can significantly vary depending on the specific problem and its representation within the Hamiltonian of the quantum system. While some problems may naturally exhibit a polynomially decreasing gap, others may have an exponentially shrinking gap, rendering them unsuitable for efficient quantum annealing. Notably, the well-known NP-Hard problem, 3SAT, has been shown to present an exponentially decreasing energy gap, thus categorizing it among the less favorable problems for quantum annealing. In

addition, for many practical problems, the behavior of their minimum energy gap is unknown beforehand, making it challenging to schedule the annealing process effectively and validate the results.

As of 2023, quantum annealing hardware, such as systems produced by D-Wave, typically operates on stoquastic Hamiltonians. A stoquastic Hamiltonian is one where all the off-diagonal elements in the standard computational basis are real and non-positive, which leads to certain computational simplifications. This limitation constrains the types of problems that quantum annealing can effectively address.

11.3 Quantum Bomb and Quantum Money

In this section, we explore the diversity of quantum algorithms through the study of two distinctive examples: the quantum bomb test algorithm and quantum money. These algorithms, while less commonly cited than DJ (Deutsch-Jozsa) and BV (Bernstein-Vazirani), Simon, or Shor, demonstrate the breadth and versatility of quantum computing. They elucidate a variety of intricate aspects, including quantum measurements, probabilistic computational methods, and post-quantum cryptographic techniques.

11.3.1 The Quantum Bomb Test Algorithm

This algorithm was initially conceptualized in a 1993 thought experiment by A. Elitzur and L. Vaidman[40]. It provides a compelling showcase for the unique principles of quantum measurements, distinguishing them sharply from their classical counterparts. This algorithm is also a cornerstone in the study of interaction-free measurements in quantum mechanics.

1 The Scenario

Consider a situation where you are presented with an assortment of indistinguishable boxes. While some contain quantum bombs, others are empty or ‘duds’. The quantum bombs are sensitive to a single photon in horizontal polarization. The task at hand is to ascertain whether a given box contains a bomb without detonating it.

In the classical realm, any attempt to test the box’s contents would result in a conundrum: if the box does not explode, you cannot conclusively determine whether it contains a bomb. If it explodes, you obviously know it was a bomb, but the object is destroyed. Essentially, it is a mission impossible in the classical realm.

2 Measuring in the Computational Basis

To cast the problem within the quantum measurement framework, let’s represent a dud by an identity gate I . A measurement of an input state $|0\rangle$ will result in the output state $|0\rangle$, and similarly for $|1\rangle$.

A bomb, on the other hand, acts as a measuring device that triggers if it encounters the input state $|1\rangle$. Although not a unitary transformation, such a device can be physically implemented without involving an actual bomb. For instance, a polarization splitter in photonics could execute this conditional action.

3 Measuring in the $\{|+\rangle, |-\rangle\}$ Basis

When considering quantum properties, the scenario changes drastically. By introduc-

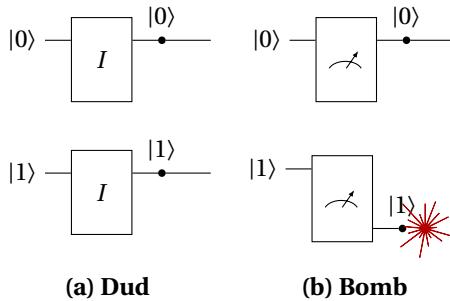


Figure 11.11: Quantum Bomb Tester - Classical Scenario

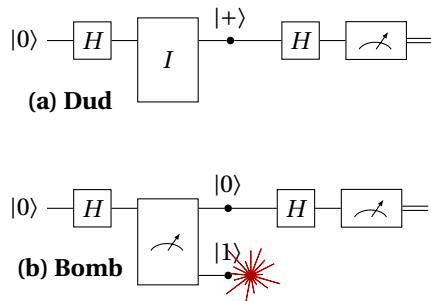


Figure 11.12: Quantum Bomb Tester - Measuring in the $\{|+\rangle, |-\rangle\}$ Basis

ing the bomb to a superposition state and measuring it in a different basis, we open up new possibilities. In particular, we can initialize our circuit with $|0\rangle$ and use two Hadamard gates to sandwich either the dud or the bomb. The initial Hadamard gate transforms $|0\rangle$ into $|+\rangle$. The subsequent Hadamard gate changes the measurement basis to $\{|+\rangle, |-\rangle\}$.

In this setup, the bomb has a 50% probability of exploding and a 50% chance of remaining intact. If it doesn't explode, a subsequent measurement will yield either $|0\rangle$ or $|1\rangle$ with equal probability. For the dud, the output remains as the input, invariably resulting in the measurement outcome $|0\rangle$.

Thus, there exists a net 25% chance of identifying a bomb without causing it to detonate. Even though not perfect, a significant improvement over the classical scenario!

4 The Quantum Bomb Test Algorithm

We now elaborate on an algorithm designed to detect the presence of a quantum bomb with an arbitrarily low probability of detonation. This algorithm is encapsulated in the quantum circuit depicted in Fig. 11.13.

The input state employed for this algorithm is as follows:

$$|\varepsilon\rangle = \cos \varepsilon |0\rangle + \sin \varepsilon |1\rangle, \quad (11.39)$$

where ε is a small angle defined by $\varepsilon = \frac{\pi}{2N}$, with N representing a large integer.

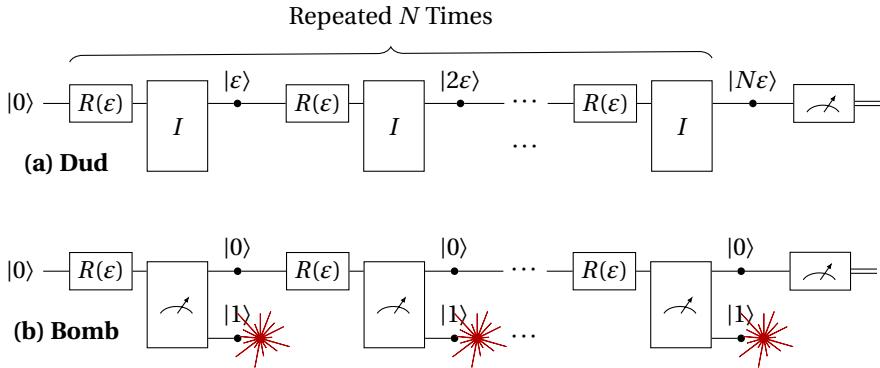


Figure 11.13: The Quantum Bomb Test Algorithm

To achieve this input state, $|\varepsilon\rangle$, we apply the $R(\varepsilon)$ gate to an initial state of $|0\rangle$:

$$R(\varepsilon) = \begin{bmatrix} \cos \varepsilon & -\sin \varepsilon \\ \sin \varepsilon & \cos \varepsilon \end{bmatrix}. \quad (11.40)$$

In the case of a dud, the qubit evolves as $R(\varepsilon)^N |0\rangle = \cos(N\varepsilon) |0\rangle + \sin(N\varepsilon) |1\rangle$ before measurement. The outcome will yield $|1\rangle$ with a probability of $\sin^2(N\varepsilon) = 1$, correctly identifying the dud.

In the case of a bomb, upon measurement, the probability of receiving $|1\rangle$ and thus triggering the bomb is $\sin^2(\varepsilon) \approx \varepsilon^2$, as per the small-angle approximation. If this does not occur, the qubit will collapse to $|0\rangle$, and the circuit will iterate.

After N iterations, the probability of triggering the bomb by that point is approximately $\frac{\pi^2}{4N}$, which can be made arbitrarily low with a large enough N . Conversely, the likelihood of identifying the bomb without triggering it, given by $\cos^{2N}(\varepsilon)$, is approximately $1 - \frac{\pi^2}{4N}$, and can be made arbitrarily close to 1.

Key Takeaways

The Quantum Bomb Test Algorithm elucidates several key concepts:

- Similar to certain classical probabilistic algorithms, quantum algorithms can be designed to perform specific tasks with arbitrary precision, involving tradeoffs in resource usage or running time, albeit never achieving absolute certainty.
- While quantum algorithms are typically conceptualized as a sequence of unitary gates culminating in measurements—a concept that might be referred to as the unitary computing model—they may also benefit from incorporating intermediate non-unitary *branching* devices.
- Interaction-free quantum measurements, allowing for the detection or inference of an object's presence without directly interacting with it, can be conducted with arbitrary precision, showcasing a unique aspect of quantum measurements that has no direct analogue in classical physics.

5 Further Information

For further developments and in-depth analysis of this subject, the reader is referred to the work on quantum measurement detection algorithms by Lugilde Fernandez et al. [68].

11.3.2 *Basics of Quantum Money $|\$ \rangle$

Quantum money employs quantum mechanical principles to create coins (or banknotes) that are inherently secure against counterfeiting. This notion was pioneered by Wiesner in 1983 and has since seen extensions and variations by subsequent researchers.

Unlike traditional digital currencies, which derive their security from computationally hard problems, quantum money exploits unique features of quantum mechanics. In particular, the no-cloning theorem establishes that an arbitrary unknown quantum state cannot be perfectly cloned. This quantum feature offers a security paradigm that transcends what is feasible with classical approaches. Nonetheless, the development of publicly verifiable quantum money remains an elusive, open challenge.

Two primary types of quantum money exist: private-key and public-key. Each has its own set of advantages, disadvantages, and appropriate use-cases, with both types leveraging quantum mechanics for enhanced security.



While digital currencies like Bitcoin are well-established in the financial world, quantum money is currently an active area of research, particularly significant in the field of quantum cryptography, exploring new frontiers in secure transactions.

1 Key Concepts

Money State

The money state, denoted as $|\$ \rangle$, is a multi-qubit state embodying a quantum coin. It is designed to resist cloning, leveraging the no-cloning theorem. It encodes quantum information critical for the authentication and verification of the value of quantum money.

$$s = \dots \dots$$

$$|\$_s\rangle = \bigotimes_{i=1}^n |\psi_i\rangle$$

A Quantum Coin
($s, |\$ \rangle$)

The number of qubits in $|\$ \rangle$, denoted n , is chosen as a tradeoff between security and efficiency, e.g., 256.

Serial Number

The serial number, represented as s , is a classical identifier uniquely associated with each money state. While the serial number is publicly readable, the corresponding quantum state, which constitutes the money, is usually secret and challenging to replicate. In various quantum money protocols, the serial number is employed to query information about the money state and is not directly used in the verification and authentication processes.

Private Key

The private key is confidential information held by the issuing bank (or mint), essential for authenticating or verifying a quantum money state. In private-key

schemes, the private key allows for the validation of the money state's authenticity when correlated with its serial number. In public-key schemes, the private key is involved in generating the serial number and money state. In all schemes, the bank rigorously protects the private key to prevent unauthorized access and potential counterfeiting.

Public Key

The public key is disseminated openly, enabling the public to authenticate a quantum money state. Public-key schemes enable decentralized verification without requiring the bank's involvement in each transaction. Despite its public accessibility, the public key is designed not to reveal enough information to counterfeit the money state.

2 Private-Key Quantum Money

Private-key quantum money schemes necessitate the involvement of the issuing bank for authentication, ensuring centralized control. Here, the no-cloning theorem guarantees the integrity and uniqueness of the quantum money state. The basic protocol is outlined below:

1. Money State Generation: The bank generates a serial number s and generates a corresponding private key K_s that remains confidential. Based on the private key, the bank prepares a quantum money state, $|\$_s\rangle$.
2. Issuance to Customer: The bank transfers $|\$_s\rangle$ and its serial number s to a customer.
3. Centralized Verification: The bank alone can validate the quantum money state using the private key K_s , which contains the required measurement angles for measurement.
4. Transaction Confirmation: Upon successful verification, the transaction is approved.

Private-key quantum money offers a more straightforward implementation but introduces scalability issues, as the bank needs to be part of every transaction, similar to a credit card payment. The scheme also shares challenges common to public-key systems, such as the need for quantum memories and error correction.

3 The Wiesner Quantum Money

The Wiesner quantum money, introduced in 1983[90], is a private-key scheme that utilizes a simple yet effective approach for secure quantum currency. In this scheme, each quantum coin, represented by $(s, |\$_s\rangle)$, is associated with a corresponding private key $K_s \in \{0, 1, +, -\}^n$. The subset $\{0, 1\}$ corresponds to the computational basis states $\{|0\rangle, |1\rangle\}$ in $|\$_s\rangle$, and $\{+, -\}$ corresponds to the Hadamard basis states, $\{|+\rangle, |-\rangle\}$. For example, $K_s = (0, +, 1, 0, -, +, \dots)$ corresponds to $|\$_s\rangle = |0 + 10 - + \dots\rangle$.

To verify a coin, the bank measures each qubit of $|\$_s\rangle$ in the appropriate basis using K_s , checking for the correct outcome. This process ensures the integrity of the authentic coin's state $|\$_s\rangle$. Conversely, measurements against an incorrect K_s on a counterfeit coin will alter $|\$_s\rangle$ due to state collapse, rendering replication or verification without K_s impossible.

In the original scheme, the issuing bank maintained a database of all (s, K_s) pairs. Modern adaptations often derive K_s from s using a secure hash function combined with a secret key (or salt) held by the bank, thereby eliminating the need for such a database. However, this approach introduces potential security concerns associated with the classical hash function, as it is not unconditionally secure.

The Wiesner scheme bears resemblance to the BB84 QKD protocol discussed in § 5.5. In fact, it served as an inspiration for the development of BB84 in 1984.

It is fascinating to examine how quantum money can be forged (which we refer to as attacks) and the possible countermeasures. Through examining these, we can gain significant insight into the issues and protocols of quantum security.

Measure and Replicate Attack

Suppose a counterfeiter measures each qubit of a quantum coin in the computational basis. Since the money state consists of $\{|0\rangle, |1\rangle\}$ with a 50% probability, and $\{|+\rangle, |-\rangle\}$ with a 50% probability, the results would be correct for half of the qubits.

The counterfeiter can then duplicate as many coins as he likes. But what is the chance for his forged coin to pass authentication at the bank? For each qubit, if the correct basis is indeed $\{|0\rangle, |1\rangle\}$, he gets a PASS, which has a probability of $\frac{1}{2}$. If the correct basis happens to be $\{|+\rangle, |-\rangle\}$, measuring a computational basis state in the Hadamard basis has a $\frac{1}{2}$ chance to obtain the authentic result. So the overall passing probability for each qubit is $\frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$. For $n = 256$ qubits, the overall passing probability is $(\frac{3}{4})^n \cdot 10^{-32}$. Therefore, the counterfeiter has to replicate approximately 10^{32} coins for this simple attack to be profitable, which is deemed impractical.

A possible countermeasure for the bank is limiting the rate of authentication trials per serial number, say to 1 per second. Note the bank should not simply blacklist a serial number after seeing a bad authentication, because doing so would allow a bad actor to cancel another person's quantum money.

Qubit-by-Qubit Learning Attack

Suppose the bank returns a quantum coin to the customer even if the authentication FAILs. In this scenario, a counterfeiter can devise an attack by learning the qubits one by one [69]. The attacker modifies the state of a single qubit (the i -th) in a legitimate coin to, say, $|0\rangle$, and sends this tampered coin to the bank for authentication. If the correct state should be $|1\rangle$, the coin will be rejected with a probability of 100%. If the correct state is $|+\rangle$ or $|-\rangle$, there is a $\frac{1}{2}$ chance of the coin being rejected. Therefore, the probability for a coin with an incorrect qubit state to be rejected is $\frac{1}{3} + \frac{2}{3} \cdot \frac{1}{2} = \frac{2}{3}$, and $\frac{1}{3}$ for it to not be rejected.

Since the authentication process does not alter the other qubits, which are in the eigenstates of the measurement bases, the counterfeiter can repeat this process. After m trials, the probability for a qubit with the wrong state to be rejected by the bank is at least $1 - (\frac{1}{3})^m$. For example, with $m = 10$ trials, the attacker can determine with high probability if the new state of the coin is incorrect. By systematically testing all four states, $|0\rangle, |1\rangle, |+\rangle$, and $|-\rangle$, the attacker can ascertain the state of the i -th qubit. This process can be repeated to learn the states of all the qubits, requiring a total of mn trials. As the number of trials is linear with n , this strategy represents a highly practical attack.

From this analysis, it is evident that to prevent such attacks, the bank should not return a failed authentication coin to the customer.

The Adaptive Attack

Suppose the bank destroys any coin (or blacklists its serial number) if authentication FAILs, but returns the coin to its customer if authentication PASSes. A clever attack exploiting this situation is derived from the quantum bomb test algorithm (refer to § 11.3.1)[31]. Here, an authentication-FAIL (resulting in the coin's destruction) is analogous to the bomb in the quantum bomb test. The coin undergoes N measurements for each test, but the total probability of triggering an authentication failure and destroying the coin is capped at $\frac{\pi^2}{4N}$, which can be made arbitrarily small for a sufficiently large N .

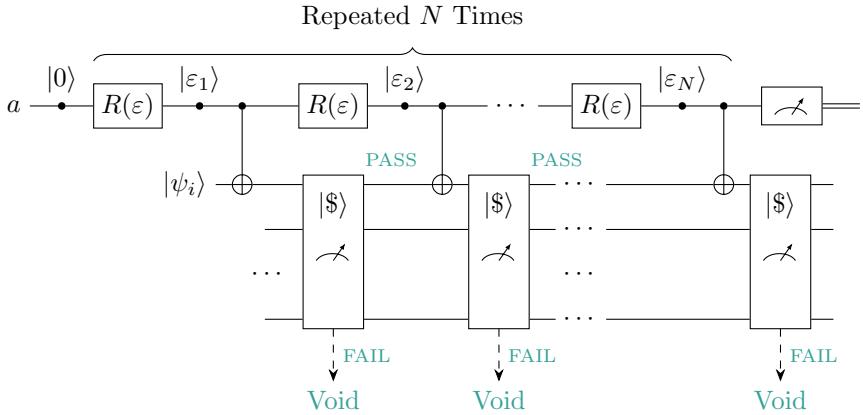


Figure 11.14: Adaptive Attack on Wiesner Quantum Money

The attack scheme, illustrated in Fig. 11.14, assumes the attacker can attach an ancilla qubit a to the money state $|\$\rangle$ and apply a quantum gate (e.g., CNOT) between it and any qubit $|\psi_i\rangle$ in $|\$\rangle$. Before each authentication, the ancilla qubit is rotated by a small angle $\varepsilon = \frac{\pi}{2N}$ and then CNOT'ed with $|\psi_i\rangle$. The state of the ancilla qubit just before the first CNOT gate is:

$$|\varepsilon_1\rangle = \cos \varepsilon |0\rangle + \sin \varepsilon |1\rangle. \quad (11.41)$$

After the CNOT gate, the combined state of the ancilla and the i -th money qubit is:

$$|\psi_{\text{combined}}\rangle = \text{CNOT}(|\varepsilon_1\rangle \otimes |\psi_i\rangle) \quad (11.42a)$$

$$= \begin{cases} \cos \varepsilon |00\rangle + \sin \varepsilon |11\rangle & \text{for } |\psi_i\rangle = |0\rangle, \\ \cos \varepsilon |01\rangle + \sin \varepsilon |10\rangle & \text{for } |\psi_i\rangle = |1\rangle, \\ (\cos \varepsilon |0\rangle + \sin \varepsilon |1\rangle) \otimes |+\rangle & \text{for } |\psi_i\rangle = |+\rangle, \\ (\cos \varepsilon |0\rangle - \sin \varepsilon |1\rangle) \otimes |-\rangle & \text{for } |\psi_i\rangle = |-\rangle. \end{cases} \quad (11.42b)$$

Exercise 11.15 Verify Eq. 11.42.

The probability of the bank's subsequent authentication measurement yielding a FAIL is:

$$p(\text{FAIL}) = \begin{cases} \sin^2 \varepsilon \approx (\frac{\pi}{2N})^2 & \text{for } |\psi_i\rangle = |0\rangle, \\ \sin^2 \varepsilon \approx (\frac{\pi}{2N})^2 & \text{for } |\psi_i\rangle = |1\rangle, \\ 0 & \text{for } |\psi_i\rangle = |+\rangle, \\ 0 & \text{for } |\psi_i\rangle = |-\rangle. \end{cases} \quad (11.43)$$

It becomes apparent that $|\psi_i\rangle = |+\rangle$ or $|-\rangle$ corresponds to a dud, while $|\psi_i\rangle = |0\rangle$ or $|1\rangle$ corresponds to a bomb. If the authentication succeeds, which is highly probable, the money state reverts to $|\psi_i\rangle$. For $|\psi_i\rangle = |0\rangle$ or $|1\rangle$, the ancilla and coin qubit were entangled, causing the ancilla to collapse back to $|0\rangle$. For $|\psi_i\rangle = |+\rangle$, the ancilla remains unchanged, while for $|\psi_i\rangle = |-\rangle$, it changes to $\cos \varepsilon |0\rangle - \sin \varepsilon |1\rangle$.

If the coin survives m measurements, the ancilla state becomes:

$$|\varepsilon_m\rangle = \begin{cases} |0\rangle & \text{for } |\psi_i\rangle = |0\rangle, \\ |0\rangle & \text{for } |\psi_i\rangle = |1\rangle, \\ \cos m\varepsilon |0\rangle + \sin m\varepsilon |1\rangle & \text{for } |\psi_i\rangle = |+\rangle, \\ \cos \varepsilon |0\rangle - \sin \varepsilon |1\rangle & \text{for } |\psi_i\rangle = |-\rangle \text{ and odd } m, \\ |0\rangle & \text{for } |\psi_i\rangle = |-\rangle \text{ and even } m. \end{cases} \quad (11.44)$$

After N measurements, assuming N is even and $N\varepsilon = \frac{\pi}{2}$, the ancilla state resolves to:

$$|\varepsilon_N\rangle = \begin{cases} |0\rangle & \text{for } |\psi_i\rangle = |0\rangle, |1\rangle, \text{ or } |-\rangle, \\ |1\rangle & \text{for } |\psi_i\rangle = |+\rangle. \end{cases} \quad (11.45)$$

At this juncture, the attacker measures the ancilla qubit. If the result is $|1\rangle$, it indicates that the i -th money qubit is definitively $|+\rangle$. However, the attacker cannot discern whether the correct basis is $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$. To determine this, the attacker replaces the CNOT gate with a controlled-negative-NOT gate (refer to § 7.1.3), thus discerning if the qubit is in the $|-\rangle$ state. With a total of $2nN$ trials, the attacker has a high probability, approximately $1 - \frac{n\pi^2}{2N}$, of learning the correct measurement bases without detection. This strategy enables the measurement of each qubit in its correct basis to deduce the entire money state.

This attack demonstrates that Wiesner's money scheme can only be secure if the bank replaces validated coins.

i

If an attacker can replicate a quantum coin with a high probability, does this violate the no-cloning theorem? No, because the theorem pertains to the deterministic replication of an arbitrary quantum state.

Exercise 11.16 Work out the attack procedure for the adaptive attack on Wiesner's coin using controlled-Z and controlled-negative-Z gates.

4 Challenge-and-Response Authentication

Private-key quantum money protocols, akin to modern credit card systems, require centralized authentication, imposing significant demands on quantum networking and communication infrastructure. To mitigate these challenges, several alternative protocols have been proposed, including challenge-and-response and public-key systems.

Consider Wiesner's quantum money for illustrating the challenge-and-response authentication process. It initiates with the coin owner transmitting the serial number s to the bank via a *classical*, potentially insecure, channel. The bank then generates a *random* challenge string $C_s \in \{0, 1, +, -\}^n$. This string, formatted like the private key K_s , shares a 50% overlap with it. The owner measures the money state $|\$\rangle$ with respect to C_s and relays the outcomes back to the bank. The bank then authenticates the coin by comparing these measurements with the recorded states of the overlapping qubits.

$$s = \dots \dots$$

$|\$_{s,1}\rangle = \bigotimes_{i=1}^n |\psi_i\rangle |1\rangle$

 \dots

 $|\$_{s,m}\rangle = \bigotimes_{i=1}^n |\psi_i\rangle |m\rangle$

A Quantum Coin
with Multiple $|\$\rangle$

This approach is reminiscent of the key-sifting protocol in the BB84 quantum key distribution (QKD) method (see § 5.5). A significant limitation for quantum money using this approach, however, is the degradation of qubits with each authentication attempt. Specifically, half of the qubits in the coin are compromised during the process. To counter this, quantum coins could be designed to associate multiple identical money states with a single serial number, while the bank maintains records of the invalidated states.

Informationally Secure vs. Computationally Secure

This challenge-and-response authentication protocol, as well as the BB84 and E91 QKD protocols, are informationally secure.

Informationally secure systems guarantee security based on the principles of information theory, ensuring that a cryptographic system remains secure regardless of the computational power available to an adversary. This means that even with unlimited computing resources, an attacker cannot gain any useful information. The one-time pad is a classic example of an informationally secure system.

In contrast, computationally secure systems rely on the assumption of limited computational resources. Their security is based on the difficulty of solving certain mathematical problems, like factoring large numbers, within a practical amount of time or resources. Most modern cryptographic systems, such as RSA and ECC, are computationally secure. Their security could potentially be compromised by future advancements in computing power.

5 Public-Key Quantum Money

Basic Protocol

The essence of public-key quantum money lies in creating a money state that can be efficiently produced by the bank (mint) and verified by the merchant, yet is

infeasible to duplicate under both quantum and classical computational security frameworks.

This system combines quantum computing with classical cryptography to enable a secure and decentralized protocol. The fundamental steps are:

1. Money State Generation: The bank generates a quantum coin (banknote) with a distinct serial number s , a quantum money state $|\$_s\rangle$, and a corresponding verification protocol P_s . These components must satisfy:
 - (a) The quantum state $|\$_s\rangle$ consistently passes the verification test defined by P_s .
 - (b) The verification protocol P_s is non-destructive, preserving the state $|\$_s\rangle$ post-verification.
 - (c) It is computationally infeasible for a potential counterfeiter, even with access to both $|\$_s\rangle$ and P_s , to replicate the quantum state such that both originals and duplicates pass P_s .

The bank publishes a registry of valid serial numbers $\{s\}$ along with their associated verification protocols $\{P_s\}$.

2. Issuance to Customer: The bank issues the quantum coin $|\$_s\rangle$, along with its serial number s , to a customer.
3. Public Verification: The authenticity of $|\$_s\rangle$ can be verified by anyone through two criteria: (1) confirming that the serial number s is publicly listed and (2) validating that $|\$_s\rangle$ satisfies the publicly accessible protocol P_s .
4. Transaction Confirmation: The quantum money is accepted by the recipient, and the transaction is deemed complete, if both criteria are successfully met.

While conceptually appealing, the practical implementation of public-key quantum money faces hurdles, such as the need for durable quantum memories and efficient, error-tolerant quantum operations. Furthermore, developing a practical realization of the tuple $(s, |\$_s\rangle, P_s)$ that fulfills all the stated requirements, even under computational security, remains a formidable challenge.

A Basic Implementation Scheme

Implementing public-key quantum money is an intricate task, involving quantum computing, advanced mathematics, and cryptography. We present a simplified, abstract protocol to highlight key concepts and challenges. Each step here is non-trivial:

1. Set Partition: Begin with a vast set of quantum states, such as:

$$B = \{|i\rangle\}, \quad \text{with } i \in [1, 2^n], \quad \text{where } n = 1024. \quad (11.46)$$

Choose a large subset T of B , but significantly smaller than B , for instance,

$$T \subset B, \quad |T| = 2^{\frac{n}{2}}. \quad (11.47)$$

Define a many-to-one function f :

$$f : B \mapsto T, \quad (11.48)$$

such that $f^{-1}(t)$ has a substantial number of preimages for any $t \in T$, say,

$$N \equiv \frac{|B|}{|T|} = 2^{\frac{n}{2}}. \quad (11.49)$$

2. Oneway Function: Ensure that computing $f(b)$ is feasible in polynomial time, whereas finding $f^{-1}(t)$ is infeasible, even with quantum resources, for large n .
3. Money Generation: Under the aforementioned conditions, the bank generates money states and serial numbers as follows:
 - a. Create a uniform superposition of all states in B :

$$|\Psi_0\rangle = \frac{1}{\sqrt{|B|}} \sum_{b \in B} |b\rangle |0\rangle, \quad (11.50)$$

where $|0\rangle$ is an $\frac{n}{2}$ -qubit register initialized to zero.

- b. Compute f into the second register:

$$|\Psi_1\rangle = \frac{1}{\sqrt{|B|}} \sum_{b \in B} |b\rangle |f(b)\rangle. \quad (11.51)$$

- c. Measure the second register to obtain a value s , collapsing the state to (see § 6.4.3 for a detailed explanation):

$$|\Psi_2\rangle = \frac{1}{\sqrt{N}} \sum_{\substack{b \in B \\ f(b)=s}} |b\rangle |s\rangle. \quad (11.52)$$

The random measurement outcome s from $\{f(b)\}$ can be used as the serial number, with the first register of $|\Psi_2\rangle$ as the money state $|\$_s\rangle$:

$$|\$_s\rangle = \frac{1}{\sqrt{N}} \sum_{\substack{b \in B \\ f(b)=s}} |b\rangle. \quad (11.53)$$

Since $f^{-1}(t)$ is assumed to be computationally hard, $|\$_s\rangle$ is infeasible to compute from s .

4. Public Verification: An authentic $|\$_s\rangle$ is a uniform superposition of all $|b\rangle$'s for which $f(b) = s$. Verification involves two key aspects:
 - a. Alignment with the Serial Number: The money state $|\$_s\rangle$ should be such that when f is applied to each $|b\rangle$ in the superposition, the output consistently equals s . This can be implemented as an eigenvalue relationship $\hat{f}(|\$_{s,1}\rangle) = s |\$_{s,1}\rangle$, where \hat{f} is the quantum version of f .
 - b. Fidelity of the Superposition: In the scheme presented here, $|\$_s\rangle$ must be a full, uniform superposition of all $|b\rangle$'s such that $f(b) = s$, not just a single state or a different superposition.

The bank consolidates this verification method into a verification operator P_s , and publishes its formula along with the serial number s . The verification protocol must be implemented such that replicating the quantum state $|\$_s\rangle$ is computationally infeasible, even with access to both s and P_s .

For instance, knot-invariance based quantum money, as in [41], utilizes Alexander polynomials for the function f . It employs a superposition of all possible Reidemeister moves — elementary transformations preserving knot invariance — to ensure that $|\$_s\rangle$ is a full, uniform superposition.

6 * Current Development

Quantum Money Protocols

A variety of quantum money protocols have been proposed, with many experiencing subsequent cryptographic breaks. Key contributions in this area (as of 2023) include:

- Farhi, Gosset, Hassidim, Lutomirski, Shor, 2009: based on knot invariants [41].
- Aaronson and Christiano, 2012: predicated on subspaces [12] (proven broken).
- Mark Zhandry, 2017: rooted in complexity-theoretic foundations [96] (proven broken).
- Kane, Sharif, Silverberg, 2018, 2021: utilizing quaternion algebra [57].
- Shor, Khesin, Lu, 2021, 2022: founded on lattice cryptography [59] (proven broken).

For a more comprehensive understanding, consult the following works:

- Another round of breaking and making quantum money [67].
- Cryptanalysis of three quantum money schemes [24].

Quantum Lightning

Quantum lightning [96] is a concept related to quantum money, representing an even stronger version of it. In a quantum money scheme, the bank (or mint) issues quantum banknotes that are hard to counterfeit due to their quantum properties. The mint is the only entity that can produce these valid banknotes, and anyone can verify their authenticity.

In contrast, quantum lightning takes this concept further by ensuring that not even the mint can duplicate the banknotes once they are issued. Each quantum banknote in a quantum lightning scheme is unique and cannot be reproduced, not even by the issuer. This feature offers a higher level of security and uniqueness for each banknote.

Quantum lightning is still a theoretical concept and part of ongoing research in quantum cryptography. It presents significant challenges in terms of practical implementation and security proof.

Post-Quantum Cryptography

Cryptosystems fundamentally rely on mathematically challenging problems. Such problems vary widely and lead to a diverse set of cryptosystems, including RSA, discrete logarithm, lattices, multivariate equations, knapsack problems, decoding challenges, among others.

Although most of these schemes are characterized by lengthy public keys, RSA and discrete logarithm based systems have gained popularity primarily due to their shorter key lengths. This advantage is often attributable to the structured mathematics underlying these schemes. Ironically, this very structure renders them vulnerable to quantum attacks, particularly through the application of Shor's algorithm.

In light of this, lattice-based cryptography has emerged as a frontrunner in post-quantum cryptographic research. It offers a balanced combination of security, efficiency, and speed.

For additional resources on this subject, refer to:

- Post-quantum cryptosystems for Internet-of-Things: A survey on lattice-based algorithms [14].
- The science and information organization [42].
- Evaluation and comparison of lattice-based cryptosystems for a secure quantum computing era [77].

Significance of the Field

The advent of quantum computing has cast a dual shadow on the field of cryptography. On one hand, Shor's algorithm poses a significant challenge by threatening the integrity of classical encryption schemes through efficient integer factorization. On the other hand, quantum computing furnishes a suite of novel techniques that bolster encryption and security protocols. This positions quantum computing not merely as a disruptive force but also as an enabling technology. A prime example of such an innovation is quantum money.

Quantum money transcends its role as a singular cryptographic construct and serves as an elemental foundation for various unclonable cryptographic architectures. It underpins more elaborate protocols like quantum copy protection and chainless blockchain systems. Consequently, mastering the subtleties of quantum money serves as a seminal vehicle for a comprehensive understanding of quantum cryptography at large.

The study of quantum money bears significance that extends well beyond the conventional bounds of encryption and security. It functions as a versatile instrument for securing a wide array of digital assets, rendering them virtually induplicable. This places quantum money at the confluence of quantum computing, cryptography, and information theory, and opens up new horizons for secure, efficient, and verifiable transactions in the quantum epoch.

Such multifaceted relevance elevates the subject of quantum money from a point of theoretical interest to a pivotal research area that has the potential to reshape the landscape of secure transactions and information exchange in the future.

11.4 Summary and Conclusions

Foundational Algorithms

The central focus of this chapter lies in offering a panorama of quantum algorithms. We began by delving into the foundational quantum algorithms, exemplified by the Deutsch-Jozsa Algorithm. While its practical applications may be limited, the algorithm serves as a seminal touchstone, elucidating the potential computational advantages of quantum systems over classical ones. We scrutinized its fundamental principles and proposed a generalized version of the algorithm.

NISQ/Utility Hybrid Algorithms

Transitioning from foundational algorithms, we examined the state of the art in quantum computing as of early 2024, spanning from NISQ devices to the nascent era of quantum utility. We focused on Quantum-Classical Hybrid Algorithms, highlighting their role in navigating the limitations of NISQ technologies and the emerging utility-era systems.

Optimization in the NISQ/Utility Era

The chapter then transitioned to explore algorithms particularly relevant to the NISQ era, focusing on solving the Quadratic Unconstrained Binary Optimization (QUBO) problem. We introduced the Variational Quantum Eigensolver (VQE) and the Quantum Approximate Optimization Algorithm (QAOA), both tailored to operate within the constraints of current-day quantum devices. By employing these algorithms, we attempted to solve problems of practical significance, like the Max-Cut problem, thereby demonstrating their utility in real-world applications.

Innovative Quantum Algorithmic Concepts

The chapter concluded with an exploration of innovative quantum algorithms, such as the Quantum Bomb Measurement and Quantum Money. These algorithms not only push computational boundaries but also introduce new aspects like intricate quantum measurements and post-quantum cryptographic techniques.

Pedagogical Aims

Throughout this chapter, our goal has been to provide a comprehensive, albeit not exhaustive, overview of the quantum algorithm landscape. By selecting representative algorithms from different categories, we aimed to illustrate the past milestones, current challenges, and future potentials in the realm of quantum computing.

Upcoming Topics

In the ensuing chapter, we will delve into the critical topic of Error Correction, focusing initially on the mathematical tool of density operators. This chapter will provide a detailed exploration of coherent vs incoherent errors and discuss the famous Shor Codes as a key technique in quantum error correction. The objective is to equip the reader with a foundational understanding of error types and mechanisms to manage them, a subject indispensable for advancing in both theoretical and practical realms of quantum computing.

Problem Set 11

- 11.1** Referring to the DJ algorithm detailed in § 11.1, for a given n , determine the number of constant functions and the number of balanced functions.
- 11.2** Assume $f(x) = \text{parity}(x)$, or equivalently, $f(x) = \text{XOR}(x)$. Show that $f(x)$ is a balanced function.
- 11.3** Validate that the circuit in Fig. 11.5 accurately represents a phase oracle for $f(x) = \overline{x_1} \oplus x_2 \oplus x_3 \oplus \overline{x_4}$.
- 11.4** The transformations detailed in Eq. 11.7 delineate the foundational mechanics of the DJ algorithm. These mechanics also underpin other quantum algorithms, including those of Simon and Shor.

Examine and elucidate the rationale underlying each transformation step detailed in Eq. 11.7. Your objective is to competently deduce and articulate the derivation of each step, utilizing the circuit representation provided in Fig. 11.1 as your foundational reference.

- 11.5** The transform below is a quantum operation used in the Deutsch-Jozsa (DJ) and Bernstein-Vazirani (BV) algorithms. It is defined as:

$$U_s |x\rangle |y\rangle = |x\rangle |y \oplus f(x)\rangle,$$

where $f(x) = x \cdot s$ is a bitwise product of x and a secret string s , and \oplus denotes bitwise XOR (or addition modulo 2).

Consider a system with n qubits for x , and an additional qubit for y . The total system size is $n + 1$ qubits.

- (a) Write out the matrix form of U_s for a 4-qubit system ($n = 3$).
 - (b) Verify that U_s is a unitary transform for any n .
 - (c) Compute the state $U_s |0^{\otimes n}\rangle |0\rangle$ for a given string s .
 - (d) Compute the state $U_s |0^{\otimes n}\rangle |1\rangle$ for a given string s .
 - (e) Compute the state $U_s |1^{\otimes n}\rangle |0\rangle$ for a given string s .
 - (f) Find a general formula for $U_s |x_1 x_2 \cdots x_n\rangle |y\rangle$ where $x_i, y \in \{0, 1\}$.
- 11.6** The **Number Partitioning Problem** (NPP) poses the following question: Given a set of N positive numbers $S = \{n_1, \dots, n_N\}$, can this set be partitioned into two disjoint subsets, such that the sums of the numbers in each subset are equal? For instance, is it possible to evenly divide a collection of assets with values n_1, \dots, n_N between two individuals?

This problem can be modeled using an Ising Hamiltonian expressed as:

$$H = \left(\sum_{i=1}^N n_i Z_i \right)^2,$$

where Z_i represents the spin variable associated with the i -th number, taking values of either $+1$ or -1 .

A solution to the NPP corresponds to a configuration of the Ising model where $H = 0$. Such a configuration ensures that the sum of the numbers n_i

corresponding to the $+1$ spins is equal to the sum of the numbers for the -1 spins. Consequently, if the ground state energy of the system is $H = 0$, the set S can be partitioned as described in the problem statement.

Your task is to create an array of N random integers and determine whether a fair partition is possible. You are to find solutions using QAOA or VQE, implemented on platforms such as Qiskit or Cirq, or via a quantum annealer. Choose a value of N that is practical for the quantum computing resources available to you, ensuring that $N > 10$ to render the problem significant.

- 11.7 The Graph Coloring Problem** (GCP) is a well-known challenge in both computer science and discrete mathematics. It entails the assignment of colors to the vertices of a graph in such a way that adjacent vertices are not colored the same, with the aim often being to minimize the total number of colors used. This problem finds practical applications in a variety of fields, including scheduling, resource allocation, frequency assignment in mobile networks, and register allocation in compiler design.

Formulating as a CUBO Problem

The GCP can be effectively formulated as a Constraint Satisfaction Problem (CSP), which is then convertible into a Combinatorial Unconstrained Binary Optimization (CUBO) problem, as detailed in § 11.2. This conversion involves several steps:

- Define Binary Variables: For a graph comprising V vertices, and with the intention to use K colors, a binary variable $x_{i,k}$ is defined for each combination of vertex i and color k . Here, $x_{i,k} = 1$ indicates that vertex i is assigned color k , and $x_{i,k} = 0$ otherwise.
- Objective Function: The CUBO formulation for the GCP prioritizes optimization in the absence of direct constraints on the binary variables. The main goal here is to minimize a particular metric or to satisfy certain conditions that are implicitly encoded within these variables. In the case of the GCP, it is crucial to encode constraints ensuring that no two adjacent vertices share the same color.
- Incorporate Constraints: Constraints are integrated into the objective function in the form of penalties. The primary constraints are as follows:

- Ensure each vertex is assigned just one color: $\sum_{k=1}^K x_{i,k} = 1$ for all i .
- Prevent adjacent vertices from sharing the same color: $x_{i,k} + x_{j,k} \leq 1$ for every edge (i, j) in the graph, applicable across all colors k .

Introducing penalty terms for these constraint violations allows the GCP to be modeled as a CUBO problem, with the objective of minimizing the total penalty to ideally achieve a value of zero when a valid coloring is found. This strategy is akin to the approach employed in solving the Traveling Salesman Problem, as elucidated in § 11.2.5.

Mapping to the Ising Model

The Ising model, as introduced in § 11.2.2, describes ferromagnetism through a mathematical framework but extends its utility to quantum computing. It

characterizes systems using discrete spin variables, which can assume states of either $+1$ or -1 , with the system's energy defined by the interactions among these spin pairs and any external magnetic fields.

Mapping the CUBO formulation of the GCP onto the Ising model involves the following procedures:

- (a) Variable Transformation: The binary variables $x_{i,k}$ are transformed into spin variables $Z_{i,k}$, following the relation $x_{i,k} = (Z_{i,k} + 1)/2$. This step converts the binary $\{0, 1\}$ representation into spin states $\{-1, +1\}$.
- (b) Energy Function: Constraints and the objective function from the CUBO model are translated into an energy function within the Ising model. Consequently, each component of the CUBO objective that relies on binary variables is mapped to a corresponding term in the Ising energy function, dependent on spin variables. The penalty terms for constraint violations manifest as interaction terms among spins.
- (c) Minimization Objective: The ultimate aim is to identify the spin configuration (the Ising model solution) that minimizes the energy function, which correlates with the optimal or a feasible solution for the original GCP.

QAOA, VQE, and quantum annealers, as discussed in §§ 11.2.4 and 11.2.6, are proficient in minimizing such Ising models, thus facilitating the resolution of GCPs on quantum computing platforms. In the simple encoding scheme explained above, referred to as the ‘one-hot encoding’, for a graph with n vertices and the goal of employing k distinct colors, a total of kn qubits are needed.

Quantum algorithms such as QAOA VQE, and quantum annealing discussed in §§ 11.2.4 and 11.2.6, are effective in minimizing Ising models that represent combinatorial optimization problems like the GCP. The one-hot encoding scheme, as described above, allows quantum computing platforms to address GCP by allocating a unique set of qubits to represent each possible color for every vertex. This approach requires a total of kn qubits for a graph with n vertices to be colored using k distinct colors. More efficient schemes, such as binary encoding for the colors, also exist.

Your Tasks

You are tasked with generating a random graph containing n vertices and exploring its colorability with 3 distinct colors. You should formulate this problem according to the guidelines provided and seek solutions employing QAOA or VQE, utilizing platforms such as Qiskit or Cirq, or through a quantum annealer. Select an n that is feasible given the constraints of the system at your disposal, yet ensure $n > 6$ for the exercise to be meaningful.

11.8 Elitzur-Vaidman Bomb-Test Algorithm The following topics build upon the core idea of the Elitzur-Vaidman bomb-tester (detailed in § 11.3.1 and § 11.3.2.3)—gathering information without direct interaction—to push the boundaries of what's possible within quantum mechanics. Researchers continue to explore these concepts not only theoretically but also experimentally, which could lead to new technologies and protocols in quantum information science.

Investigate two of the following topics, and prepare a presentation on each, discussing its concept, principles, applications, and current status.

- (a) Quantum Zeno Effect: The Quantum Zeno Effect can be seen as an extension where frequent observations prevent the evolution of a quantum state. It has been proposed that by frequently checking for the presence of the bomb (the system's state), one can effectively freeze its evolution, leading to a form of interaction-free measurement.
- (b) Counterfactual Quantum Computation: This application allows for computation to occur without running the computer in the traditional sense. It uses the principles of quantum superposition and interference to infer the result of a computation without actually performing all the steps physically.
- (c) Counterfactual Quantum Communication: This method of transmitting information without particles traveling between the sender and receiver exploits quantum entanglement and the principles behind the Elitzur-Vaidman bomb-tester to achieve communication that is effectively “interaction-free.”
- (d) Chained Quantum Zeno Effect: A series of interaction-free measurements can be chained together to create a more complex system that can perform tasks such as imaging or communication with a lower chance of interaction than the original setup.
- (e) Counterfactual Cryptography: Inspired by the bomb-tester, this extension involves secure communication where the presence of an eavesdropper can be detected without any qubits carrying the information directly from the sender to the receiver.

12. Quantum Error Correction: A Primer

Contents

12.1	Preliminary Concepts	334
12.2	* Mixed States, Density Operators, and CPTP Maps	338
12.2.1	Definitions	338
12.2.2	Interpretation and Properties	340
12.2.3	Probabilities and Expected Values	342
12.2.4	* General Single-Qubit States in the Bloch Sphere	344
12.2.5	* Unitary Evolution	346
12.2.6	* Measurements	348
12.2.7	* CPTP Maps and Quantum Channels	350
12.2.8	* Partial Trace and Reduced Density Operators	354
12.2.9	* Extension of No-Cloning and No-Communication Theorems	359
12.3	Error Mechanisms in Quantum Computing	361
12.3.1	* Coherent Errors	361
12.3.2	* Incoherent Errors	363
12.3.3	* Other Types of Errors	369
12.3.4	Performance Merit Parameters	369
12.4	Introduction to Error Correction Codes	376
12.4.1	Three-Qubit Bit-Flip Code	377
12.4.2	Three-Qubit Phase-Flip Code	380
12.4.3	* Nine-Qubit Shor Code	381
12.4.4	* Parity Measurements, Stabilizers, and Code Distance	386
12.4.5	* Further Exploration	388
12.5	Summary and Conclusions	391
	Problem Set 12	392

In this chapter, we undertake a comprehensive examination of Quantum Error Correction (QEC) in the realm of quantum computing. We commence by discussing the mathematical underpinnings of density operators, before transitioning to an exploration of both coherent and incoherent errors. The chapter culminates with

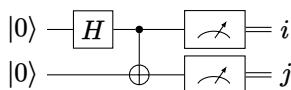
an analysis of Shor Codes, a pioneering approach in quantum error correction. The primary objective is to equip the reader with a thorough understanding of the errors that afflict quantum systems and the methodologies available to amend them. This knowledge is essential for advancing both the theoretical and practical facets of quantum computing.

12.1 Preliminary Concepts

Error correction in quantum computing is a multifaceted topic that necessitates a foundational understanding of certain core concepts.

1 The Noise Problem: An Illustrative Example

To elucidate the impact of noise, consider a quantum computer programmed to generate the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$:



In an ideal setting devoid of noise, measurements would invariably yield $ij = 00$ and 11 with equal probabilities. Contrarily, a real-world, noise-impacted system might also produce outcomes such as 01 and 10 , albeit with lower probabilities like 1%. Such deviations manifest due to noise, which disrupts the quantum state and results in aberrant measurement outcomes.

2 Decoherence, Noise, and Error: Definitions and Distinctions

The terms “decoherence,” “noise,” and “error” are often used interchangeably, yet each has a distinct meaning in quantum computing.

- **Decoherence:** Decoherence arises from the inevitable interaction between a quantum system and its surrounding environment, leading to a loss of quantum coherence (i.e., superposition). It is often modeled using density operators and is responsible for the transition of quantum states from pure to mixed.
- **Noise:** Noise encompasses all undesired transformations that quantum states may undergo during computation, stemming from various factors such as decoherence, control errors, and thermal fluctuations. Quantum channels and their associated Kraus operators are often employed to model noise.
- **Error:** An error in quantum computing refers to the deviation between the actual and expected outputs of a computation. Such errors typically stem from noise and are categorized as either coherent, which are correctable through calibration, or incoherent, which are less predictable and therefore more challenging to manage.

3 From NISQ to Fault Tolerance

The evolution of quantum computing is characterized by distinct phases, each representing a leap forward in capability and application. Currently transitioning from the Noisy Intermediate-Scale Quantum (NISQ) era, the field is moving towards an interim phase known as the “era of quantum utility,” eventually leading to the ultimate goal of full fault tolerance.

The NISQ Era

Quantum computing is currently marked by limitations in qubit count and coherence times. Amid the ongoing quest for quantum advantage, research is mainly focused on algorithmic strategies that are compatible with these NISQ constraints.

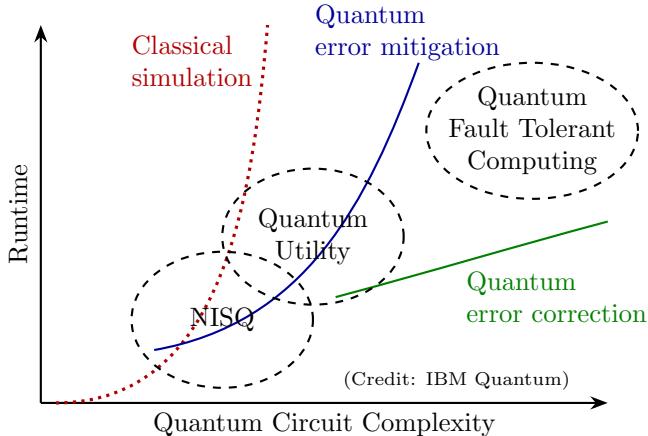


Figure 12.1: From NISQ to Fault Tolerance

The Era of Quantum Utility

As we step into the quantum utility era (circa 2024), quantum computing begins to assert its potential as a transformative technology. This phase is characterized by quantum computers not only solving problems that exceed the capabilities of brute force classical simulations but also surpassing classical systems that employ problem-specific approximation methods. This era is not merely a testament to quantum computing's theoretical prowess but also marks its evolution as an indispensable tool for scientific exploration.

The quantum utility era is distinguished by its ability to deliver reliable and substantial computations for large-scale problems, which were previously deemed infeasible without quantum intervention. With the maturation of quantum hardware, we are now witnessing quantum systems reliably handling simulation tasks involving more than 100 qubits. This milestone is underscored by the emergence of numerous “utility scale experiments,” which collectively signify a paradigm shift. These experiments represent not just a leap in quantum computational capacity but also highlight the role of quantum systems in offering novel solutions and perspectives in tackling complex problems. References such as [19, 35, 43, 62, 81, 92, 93] elucidate these advancements.

As we navigate through this era, the expectation is not just the enhancement of computational capabilities but the realization of significant, practical computational advantages. Quantum systems, in this phase, are anticipated to deliver not just solutions but also tangible utility, addressing real-world challenges and enriching the landscape of computational science.

Quantum Utility, Advantage, and Supremacy

Quantum Utility refers to the stage where quantum computers begin to solve complex, real-world problems, offering useful solutions beyond the capabilities of classical computers.

Quantum Advantage is achieved when a quantum computer outperforms classical computers for specific practical tasks, delivering solutions with superior speed, cost-effectiveness, or accuracy.

Quantum Supremacy is the threshold where a quantum computer solves a problem that is infeasible for classical computers within a reasonable time frame. While it's a significant demonstration of the potential of quantum computing, the specific problems solved may not have immediate practical applications or commercial viability.

Fault-Tolerant Quantum Computing

The Accuracy Threshold Theorem is a foundational principle in quantum error correction (QEC), suggesting that Fault-Tolerant Quantum Computing (FTQC) is attainable when the error rate per physical qubit and gate operation is suppressed below a certain critical threshold, denoted by ϵ_{th} . Achieving error rates below this threshold enables the effective logical error rate to be substantially reduced, potentially to arbitrarily low levels.

The precise threshold value, ϵ_{th} , is contingent upon the specific error-correcting code employed and the noise model assumed. Present-day quantum devices generally exhibit error rates between 10^{-2} and 10^{-3} . Recent advances in error-correcting codes and error management strategies, like the conversion of general errors into more tractable erasure errors, have fostered optimism that achieving error rates around 10^{-3} could soon render practical error correction viable. Nonetheless, further reduction of the error rate would significantly alleviate the overhead, particularly reducing the ratio of physical qubits required per logical qubit.

4 Strategies for Error Handling

Several strategies are employed to manage errors as we transition from NISQ devices to fully fault-tolerant quantum computing:

- **Error Suppression:** This approach involves modifying control signals or incorporating additional ones to suppress errors at their source, ensuring that quantum processors produce the intended outcomes more reliably.
- **Error Mitigation:** Techniques such as Zero Noise Extrapolation (ZNE) and Probabilistic Error Cancellation (PEC) aim to reduce or neutralize the impact of noise on the calculated expected values. While error mitigation has shown promise in systems with the order of 100 qubits, it faces an exponential increase in overhead as the system size grows, rendering it less feasible for very large quantum systems.
- **Error Correction:** The cornerstone of fault-tolerant quantum computing, error correction involves a set of operations and measurements designed to identify and correct errors without compromising the quantum information. The overhead for error correction scales polylogarithmically with the size of the quantum system, making it a viable strategy for scaling to larger, fault-tolerant quantum computers.

- **Error Simulation:** Simulating errors provides critical insights into the performance of quantum algorithms under realistic, noisy conditions. This strategy supports the refinement of both quantum hardware and error correction protocols, facilitating the development of more robust quantum computing systems.

5 Challenges of Quantum Error Correction

Error correction in quantum computing presents a fundamentally more complex challenge than in classical computing. This arises from a few key properties inherent to quantum mechanics:

- **Superposition:** Quantum bits (qubits) can exist in a superposition of states, in contrast to classical bits that are strictly binary. This added layer of complexity widens the scope for potential errors, as undesirable changes in the relative probabilities of the states could occur.
- **Entanglement:** The phenomenon of entanglement means that the state of one qubit can be inexplicably linked with the state of another. Consequently, an error in one qubit can reverberate through its entangled partners, making error detection and correction more intricate.
- **No-Cloning Theorem:** The no-cloning theorem (see § 5.1.3) states that it is impossible to duplicate an arbitrary unknown quantum state. This principle negates the common classical error correction practice of data copying, rendering it unviable for quantum data.
- **Measurement Disturbs Superposition:** The act of measuring a qubit disrupts its superposition, collapsing it into one of its basis states (0 or 1). The mere action of checking a qubit for an error can irreversibly alter its state, potentially eliminating the superposition necessary for computation.

6 The Path Forward

As we navigate the evolving landscape of quantum computing, several interrelated challenges emerge as central to advancing the field:

- **Quantum Error Correction and Fault Tolerance:** These remain foundational to the viability of quantum computing. Current research efforts are aimed at minimizing qubit overhead and developing innovative error-correcting codes that can more efficiently protect quantum information against errors.
- **Enhancing Quantum Coherence and Control:** Achieving extended coherence times, high-fidelity gate operations, effective decoherence control, and overall clock speed is crucial. These factors are not only vital for successful error correction but also for executing large-scale quantum algorithms. Additionally, specific error types such as coherent, calibration, state preparation, and measurement errors fall outside the purview of traditional QEC and must be addressed through other means to ensure overall system reliability.
- **Scalability and System Integration:** As we aim to increase the number of qubits, the challenges of scalability and system integration become more pronounced. This pillar focuses on the architectural and connectivity solutions required to manage an expanding quantum system effectively.
- **Quantum Software and Algorithm Development:** Parallel to hardware advancements, the creation of sophisticated quantum software and algorithms is

essential. This research avenue is dedicated to building the computational tools and frameworks necessary to harness the full potential of quantum computing.

These four aspects collectively define the critical path from the NISQ era to large-scale, practical, and fault-tolerant quantum computing.

12.2 * Mixed States, Density Operators, and CPTP Maps

To model a diverse range of quantum operations, including both pure and noisy quantum states, a more comprehensive representation beyond state vectors and unitary gates is needed. Such generalization is particularly useful for capturing the effects of noise, decoherence, or entanglement with an external environment. Density operators, also known as density matrices, and the associated mixed quantum states, serve as this general mathematical framework. This framework serves as an essential tool in quantum information and communication.

12.2.1 Definitions

1 For Pure Quantum States

Until now, we have utilized state vectors, denoted as $|\psi\rangle$, to describe quantum systems. However, these systems can also be characterized using a density operator ρ . For a quantum system isolated from its environment and in a pure state, the density operator is simply the outer product of the state vector $|\psi\rangle$ with itself. Formally, we define it as:

$$\rho = |\psi\rangle\langle\psi|. \quad (12.1)$$

■ **Example 12.1** Consider $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. The corresponding density operator is:

$$\rho = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \quad (12.2)$$

■ **Example 12.2** For $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the density operator is:

$$\rho = \begin{bmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{bmatrix}. \quad (12.3)$$

The state vectors $|\psi\rangle$ and $e^{i\phi}|\psi\rangle$, with ϕ a real number, yield identical density operators, thereby demonstrating that density operators inherently account for the physical irrelevance of the global phase in quantum state vectors.

Exercise 12.1 Show that for a pure state, $\rho^2 = \rho$, i.e., ρ is a rank-1 projection operator.

Exercise 12.2 Derive the density operator for the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.

2 For Mixed Quantum States

There are instances where the exact state of a quantum system cannot be definitively ascertained. Such a system may exist in any of n pure quantum states $|\psi_i\rangle$, each with a classical probability p_i . In these cases, the density operator provides a convenient representation:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \quad (12.4)$$

Here, the density operator for the mixed quantum state serves as a weighted average of the density operators for its constituent pure states.

■ **Example 12.3** Suppose a lab prepares an ensemble of quantum states: $|0\rangle$ with a probability of $\frac{3}{7}$ and $|1\rangle$ with a probability of $\frac{4}{7}$. This defines a *mixed state* described by:

$$\rho = \frac{3}{7} |0\rangle\langle 0| + \frac{4}{7} |1\rangle\langle 1|. \quad (12.5)$$

Note that this is completely different from the pure state $\frac{3}{7}|0\rangle + \frac{4}{7}|1\rangle$ (not normalized) or $\frac{3}{5}|0\rangle + \frac{4}{5}|1\rangle$ (normalized). ■



As demonstrated in the following exercise, n can be different from d , the dimension of the Hilbert space of $|\psi_i\rangle$. However, ρ is always a $d \times d$ matrix.

Exercise 12.3 A machine produces the state $|0\rangle$ 25% of the time, the state $|1\rangle$ 25% of the time, and the state $|+\rangle$ 50% of the time. Find the density operator of the qubit.

■ **Example 12.4 — The Uniform Mixed State.** The density operator for a mixed state with equal probabilities (50%) of being in $|0\rangle$ and $|1\rangle$ is given by:

$$\rho = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (12.6)$$

For a d -dimensional qudit, the density operator for the uniform mixed state is

$$\rho = \frac{1}{d} I_{d \times d}. \quad (12.7)$$

Remarkably, the density operator for a mixed state with 50% probability in $|+\rangle$ and 50% probability in $|-\rangle$ yields the same matrix. The same holds true for a mixed state with 50% probability in an arbitrary state $|\psi\rangle$ and 50% in its orthogonal state $|\psi_\perp\rangle$.

Hence, the density operator for the uniform mixed state is basis-invariant. This characteristic can be contrasted with the density operator for $|+\rangle$, as discussed in Example 12.1, which is not basis-invariant. ■

Exercise 12.4 Show that the mixed state with 50% probability in an arbitrary state $|\psi\rangle$ and 50% in its orthogonal state $|\psi_{\perp}\rangle$ has a density operator $\rho = \frac{1}{2} I$.

Density Operator of Density Operators

A density operator can also be expressed as a linear combination of other density operators. For example, consider two density operators ρ_1 and ρ_2 , which themselves could be representing mixed states. A new density operator can be formed as:

$$\rho' = q_1 \rho_1 + q_2 \rho_2, \quad (12.8)$$

where $q_1 + q_2 = 1$ and $q_1, q_2 \geq 0$. In this case, ρ' is a convex combination of ρ_1 and ρ_2 . Each of these density operators, ρ_1 and ρ_2 , could in turn be a weighted average of density operators for pure states. Then ρ' is also ultimately a weighted average of density operators for pure states.

3 Distinguishing Pure and Mixed States

The above example underscores an important point: although $|+\rangle$ and the uniform mixed state both yield probabilities of 50% for measurements in the computational basis, they can be distinguished in other bases. For instance, $|+\rangle$ can be identified unambiguously in the $\{|+\rangle, |-\rangle\}$ basis, whereas the uniform mixed state cannot. In general, pure states and mixed states can be distinguished in this manner.

Another criterion for distinguishing pure and mixed states is based on the value of the “purity” parameter:

$$0 < \gamma \equiv \text{tr}(\rho^2) \leq 1. \quad (12.9)$$

For a pure state, we have $\rho^2 = \rho$, leading to $\gamma = 1$. On the other hand, for a mixed state, $\gamma < 1$; for a d -dimensional uniform mixed state, γ reaches its minimum at $\frac{1}{d}$.

12.2.2 Interpretation and Properties

1 Diagonal and Off-diagonal Elements

The diagonal elements of a density operator ρ denote the probabilities of the system being observed in each of the basis states after a measurement. Specifically, the element $\rho_{ii} = \langle i | \rho | i \rangle$ gives the probability of detecting the system in the state $|i\rangle$.

The off-diagonal elements, ρ_{ij} for $i \neq j$, encapsulate the coherences between states $|i\rangle$ and $|j\rangle$, characterizing the system’s quantum superpositions. These elements may be complex, and both their magnitudes and phases are important, carrying information about the coherence of the superposition states.

The underlying reason for the above properties is demonstrated in Example 12.2 for pure states. This reasoning also applies to mixed states, as they are statistical mixtures of pure states, preserving these diagonal and off-diagonal characteristics in their density operators.

2 Diagonal Form

In its eigenbasis, a density operator ρ becomes diagonal, with eigenvalues representing the probabilities of the system residing in each eigenstate:

$$\rho = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|. \quad (12.10)$$

Each eigenstate of ρ is a pure state $|\lambda_i\rangle$. The diagonal form of ρ represents the probability distribution of the system state across these eigenstates.

This diagonal form highlights the system's probabilistic nature:

- In a pure state, ρ has exactly one diagonal element equal to 1, reflecting absolute certainty about the system's quantum state.
- For a mixed state, ρ has multiple non-zero diagonal elements, with each representing the probability of the system being found in a particular eigenstate.

Exercise 12.5 The state of a quantum system is such that it is 50% in the state $|+\rangle$ and 50% in the state $|0\rangle$. Find the diagonal form of its density operator.

3 Basic Properties

Density operators exhibit several key characteristics that confirm their suitability for physical interpretations:

1. Hermiticity: $\rho = \rho^\dagger$. This property guarantees that the eigenvalues of ρ , which are the probabilities in its eigenbasis, are real.
2. Unit Trace: $\text{tr}(\rho) = 1$. This reflects the total probability axiom, as the trace of ρ sums the probabilities for all possible outcomes of a measurement, which must equal one.
3. Positive Semi-Definiteness: For any vector $|\phi\rangle$ in the Hilbert space, $\langle\phi|\rho|\phi\rangle \geq 0$. This ensures that all eigenvalues of ρ are non-negative, thus validating that probabilities derived from ρ are always non-negative.

These properties are foundational, ensuring that density operators provide a consistent and physically plausible framework for quantum states, aligned with the postulates of quantum mechanics. Furthermore, the converse is also true: any operator (or matrix) satisfying these properties can represent a quantum state, be it pure or mixed.

Exercise 12.6 Prove the Hermitian property.

Proof. Trace One Property

$$\text{tr } \rho = \text{tr} \left(\sum_i p_i |\psi_i\rangle\langle\psi_i| \right) \quad (12.11a)$$

$$= \sum_i p_i \text{tr} (|\psi_i\rangle\langle\psi_i|) \quad (12.11b)$$

$$= \sum_i p_i = 1. \quad (12.11c)$$

□

Proof. Positive Semi-definite Property

Using the definition of ρ , we have:

$$\langle \phi | \rho | \phi \rangle = \langle \phi | \left(\sum_i p_i |\psi_i\rangle \langle \psi_i| \right) | \phi \rangle \quad (12.12a)$$

$$= \sum_i p_i \langle \phi | \psi_i \rangle \langle \psi_i | \phi \rangle \quad (12.12b)$$

$$= \sum_i p_i |\langle \phi | \psi_i \rangle|^2 \quad (12.12c)$$

$$\geq 0, \quad (12.12d)$$

because each term $p_i |\langle \phi | \psi_i \rangle|^2$ is non-negative. \square

4 Composition Property

Consider two independent systems, represented by ρ_A and ρ_B . The density operator of the composite system can be expressed as a tensor product of their individual density operators, $\rho_{AB} = \rho_A \otimes \rho_B$. A proof is outlined below:

$$\rho_A \otimes \rho_B = \sum_n \sum_{n'} p_n^A p_{n'}^B (|\psi_n^A\rangle \langle \psi_n^A|) \otimes (|\psi_{n'}^B\rangle \langle \psi_{n'}^B|) \quad (12.13a)$$

$$= \sum_n \sum_{n'} p_n^A p_{n'}^B (|\psi_n^A\rangle \otimes |\psi_{n'}^B\rangle) (\langle \psi_n^A | \otimes \langle \psi_{n'}^B |) \quad (12.13b)$$

$$= \rho_{AB}. \quad (12.13c)$$

Note that this composition property holds true only for independent systems. In the case of dependent systems, such as entangled systems, the joint density operator ρ_{AB} of a general bipartite composite system may not be expressible as the tensor product of two individual density operators.

12.2.3 Probabilities and Expected Values

The density operator provides a complete description of the quantum state, enabling the determination of outcome probabilities for any system measurement. Observable operators facilitate the extraction of measurable properties from this state, with physical quantities typically represented as statistical averages or expected values of these observables.

1 Measurement Probabilities

The probability of a system described by the density operator ρ being found in any pure state $|\psi\rangle$ is given by the expected value of the projection onto that state:

$$p = \langle \psi | \rho | \psi \rangle. \quad (12.14)$$

This equation is valid for any pure state $|\psi\rangle$, not just the basis states $\{|\psi_i\rangle\}$ in which ρ is expressed. We suggest that the reader verify this as an exercise.

Exercise 12.7 Given the density operator for a state $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$, represented by:

$$\rho = \begin{bmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{bmatrix},$$

compute the probability of measuring the system in the state $|+\rangle$ using Eq. 12.14.

* Similarity Between Mixed States

In general, there is no direct probability measure of one mixed state within another, but there are measures of similarity and distinguishability:

Fidelity

The fidelity between two density matrices ρ and ρ' is a common measure of similarity. It is defined as:

$$F(\rho, \rho') = \left(\text{tr} \sqrt{\sqrt{\rho}\rho'\sqrt{\rho}} \right)^2. \quad (12.15)$$

Fidelity equals 1 if and only if $\rho = \rho'$, and it decreases as the states become more distinguishable.

Trace Distance

Another measure is the trace distance, which quantifies the distinguishability of two quantum states. It is given by:

$$D(\rho, \rho') = \frac{1}{2} \text{tr}|\rho - \rho'|, \quad (12.16)$$

where $|A| = \sqrt{A^\dagger A}$. The trace distance ranges from 0 to 1, where 0 means the states are identical and 1 means they are completely distinguishable.

2 Expected Values of Observables

For a pure quantum state represented by $|\psi\rangle$, the expected value of an observable A is calculated as:

$$\langle A \rangle = \langle \psi | A | \psi \rangle. \quad (12.17)$$

In the case of mixed quantum states, the situation is more intricate as the system may exist in multiple states $|\psi_n\rangle$ with corresponding classical probabilities p_n . The expected value for a mixed state is essentially the weighted average of the expected values for its constituent pure states:

$$\langle A \rangle = \sum_n p_n \langle \psi_n | A | \psi_n \rangle. \quad (12.18)$$

It is noteworthy that the expected value for an observable in a mixed state can be concisely expressed using the trace operation on the density operator ρ :

$$\langle A \rangle = \text{tr}(A\rho). \quad (12.19)$$

Proof. The following steps show that $\text{tr}(A\rho) = \langle A \rangle$:

$$\mathrm{tr}(A\rho) = \mathrm{tr}\left(A \sum_n p_n |\psi_n\rangle \langle\psi_n|\right) \quad (12.20a)$$

$$= \sum_n p_n \mathrm{tr}(A |\psi_n\rangle \langle\psi_n|) \quad (12.20b)$$

$$= \sum_n p_n \langle\psi_n| A |\psi_n\rangle \quad (12.20c)$$

$$= \langle A \rangle, \quad (12.20d)$$

where we used the linearity of the trace operation in the second line and the cyclic property of the trace in the third line. \square

Exercise 12.8 For the mixed state with equal probabilities (50%) of being in $|0\rangle$ and $|1\rangle$, calculate the expected values of the Pauli observables $\langle X \rangle$, $\langle Y \rangle$, and $\langle Z \rangle$. Compare to $\langle X \rangle$, $\langle Y \rangle$, and $\langle Z \rangle$ for the pure state $|+\rangle$.

12.2.4 * General Single-Qubit States in the Bloch Sphere

1 General Single-Qubit States

Since the Pauli matrices $\{I, X, Y, Z\}$ form a complete orthogonal basis in $\mathbb{C}^{2 \times 2}$, an arbitrary single-qubit density operator can be decomposed as:

$$\rho = \frac{1}{2}(aI + xX + yY + zZ), \quad (12.21)$$

where the coefficients are given by the inner products of $\{I, X, Y, Z\}$ with ρ (defined as $A \cdot B \equiv \mathrm{tr}(A^\dagger B)$):

$$a = \mathrm{tr}(\rho I), \quad x = \mathrm{tr}(\rho X), \quad y = \mathrm{tr}(\rho Y), \quad z = \mathrm{tr}(\rho Z). \quad (12.22)$$

According to Eq. 12.19, $\mathrm{tr}(\rho A)$ equals to the expectation values of A . Therefore,

$$a = \langle I \rangle = 1, \quad x = \langle X \rangle, \quad y = \langle Y \rangle, \quad z = \langle Z \rangle. \quad (12.23)$$

Thus, we have the following important identity for single-qubit states:

$$\rho = \frac{1}{2}(I + \langle X \rangle X + \langle Y \rangle Y + \langle Z \rangle Z) \quad (12.24)$$

Exercise 12.9 The state of a quantum system is such that it is 50% in the state $|+\rangle$ and 50% in the state $|0\rangle$. Determine the Pauli representation of this quantum state.

2 Representation in the Bloch Sphere

According to Eq. 12.21, a single qubit state, pure or mixed, is uniquely identified by a vector $\mathbf{r} = (x, y, z)$, known as the *polarization vector* of the qubit. For a pure state representing a spin in the direction specified by spherical coordinate angles (θ, ϕ) , the expectation values are given by (see § 3.2.3):

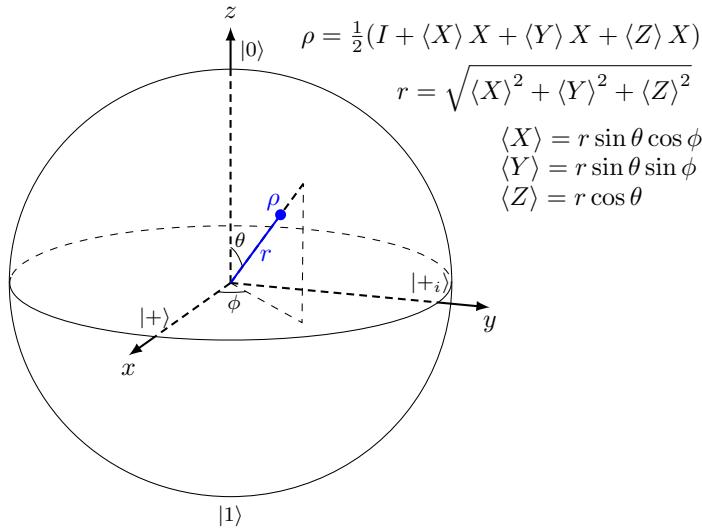


Figure 12.2: Mixed State Inside the Bloch Sphere

$$x = \sin \theta \cos \phi, y = \sin \theta \sin \phi, z = r \cos \theta. \quad (12.25)$$

Generalizing this, we can interpret \mathbf{r} geometrically as:

$$\mathbf{r} = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta). \quad (12.26)$$

where $r^2 \equiv x^2 + y^2 + z^2 \leq 1$.

Therefore, while a pure state corresponds to a point on the Bloch sphere (see § 2.4), a mixed state can be represented as a point inside the Bloch sphere, as illustrated in Fig. 12.2.

Intuitively, the length of the polarization vector (or Bloch vector), r , gives an indication of the “average” behavior of the mixed state. Specifically, r provides a measure of the “purity” of the state, varying between 0 and 1. A length of $r = 1$ corresponds to pure states, while $r < 1$ corresponds to mixed states. The closer r is to zero, the more mixed the state is.

Mathematically, Eq. 12.24, the density operator of an arbitrary qubit state, can be expressed succinctly as:

$$\rho = \frac{1}{2}(I + \mathbf{r} \cdot \boldsymbol{\sigma}), \quad (12.27)$$

and Eq. 12.23, the expectation of $\boldsymbol{\sigma}$ in this state, as:

$$\langle \boldsymbol{\sigma} \rangle = \mathbf{r}, \quad (12.28)$$

where

$$\boldsymbol{\sigma} = (\sigma_1 \equiv X, \sigma_2 \equiv Y, \sigma_3 \equiv Z). \quad (12.29)$$

Exercise 12.10 Show that for the uniform mixed qubit state (see Example 12.4), the Bloch vector is at the origin of the Bloch sphere. Therefore, the uniform mixed state is the maximally mixed qubit state.

Exercise 12.11 Explain why there is a factor $\frac{1}{2}$ in Eqs. 12.21 and 12.27.

12.2.5 * Unitary Evolution

1 Unitary Transformations

Recall (in Chapter 4) that in a closed quantum system governed by a Hamiltonian H , a pure quantum state $|\psi\rangle$ evolves according to the equation

$$|\psi\rangle \rightarrow U |\psi\rangle, \quad (12.30)$$

where the unitary operator U is defined as:

$$U = e^{-iHt/\hbar}. \quad (12.31)$$

Consequently, the density operator ρ for a mixed state transforms as:

$$\rho \rightarrow \rho' = \sum_n p_n U |\psi_n\rangle \langle \psi_n| U^\dagger. \quad (12.32)$$

This can be simplified to:

$$\rho \rightarrow \rho' = U \rho U^\dagger. \quad (12.33)$$

Exercise 12.12 The state of a quantum system is such that it is 50% in the state $|+\rangle$ and 50% in the state $|0\rangle$. An X gate is applied to the system. Using the above transformation, find the new density operator for the system.

■ **Example 12.5 — Application of Pauli Gates.** Consider the application of an X gate to a general quantum state $\rho = \frac{1}{2}(aI + xX + yY + zZ)$ in the Pauli-decomposition representation. Note that x, y, z coincide with the expected values $\langle X \rangle, \langle Y \rangle, \langle Z \rangle$ (see Eq. 12.24). The Pauli X gate, representing a π rotation around the x -axis, leaves $\langle X \rangle$ unchanged but flips the signs of $\langle Y \rangle$ and $\langle Z \rangle$. Therefore, X transforms (x, y, z) to $(x, -y, -z)$, or,

$$X \frac{1}{2}(aI + xX + yY + zZ)X = \frac{1}{2}(aI + xX - yY - zZ). \quad (12.34)$$

Extending this to a general rotation around the x -axis, $R_x(\theta)$ transforms the state (x, y, z) to $(x, y \cos \theta, z \cos \theta)$. ■

Exercise 12.13 How do the Pauli operators Y and Z , as well as the combined operator ZX , individually transform a quantum state represented by its Bloch vector $\mathbf{r} = (x, y, z)$?

2 *The von Neumann Equation

The time evolution of a density operator in a quantum system is governed by the von Neumann equation, which is a generalization of the Schrödinger equation for mixed states. For a pure state $|\psi(t)\rangle$, governed by the Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle, \quad (12.35)$$

the corresponding von Neumann equation for the density operator ρ is given by:

$$i\hbar \frac{\partial \rho}{\partial t} = [H, \rho], \quad (12.36)$$

derived as follows:

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial t} (|\psi\rangle\langle\psi|) \quad (12.37a)$$

$$= \left(\frac{\partial}{\partial t} |\psi\rangle \right) \langle\psi| + |\psi\rangle \left(\frac{\partial}{\partial t} \langle\psi| \right) \quad (12.37b)$$

$$= \frac{-i}{\hbar} H |\psi\rangle\langle\psi| + \frac{i}{\hbar} |\psi\rangle\langle\psi| H \quad (12.37c)$$

$$= \frac{-i}{\hbar} [H, \rho]. \quad (12.37d)$$

Given the linearity of quantum mechanics and density operators, this derivation extends naturally to mixed states, which are statistical ensembles of pure states, validating the von Neumann equation for mixed states as well.

If the Hamiltonian is time-independent, the von Neumann equation has the solution

$$\rho(t) = e^{-iHt/\hbar} \rho(0) e^{iHt/\hbar}, \quad (12.38)$$

which mirrors the unitary evolution of density operators (Eq. 12.33).

For a single qubit, the Hamiltonian can be decomposed in the Pauli basis, similar to Eq. 12.27:

$$H = \frac{1}{2}(H_0 I + \mathbf{H} \cdot \boldsymbol{\sigma}), \quad (12.39)$$

where $\boldsymbol{\sigma}$ is given by Eq. 12.29.

Then the von Neumann equation leads to:

$$\frac{\partial \langle \sigma_n \rangle}{\partial t} = \text{tr} \left(\sigma_n \frac{\partial \rho}{\partial t} \right) = -i \text{tr} (\sigma_n [H, \rho]) = \sum_{jk} \varepsilon_{njk} H_j \langle \sigma_k \rangle, \quad (12.40)$$

where $n, j, k \in \{1, 2, 3\}$, or

$$\frac{\partial \langle \boldsymbol{\sigma} \rangle}{\partial t} = \mathbf{H} \times \langle \boldsymbol{\sigma} \rangle. \quad (12.41)$$

12.2.6 * Measurements

Transitioning from the ket representation to density operators offers a more encompassing description of quantum states. This subsection explores how quantum measurement theory (§ 3.4) is adapted to align with the density operator framework. We will discuss the characterization of post-measurement states, the emergence of mixed states from measurements, measurements in an orthonormal basis, and the concept of POVM (Positive Operator-Valued Measure) measurements.

1 Post-Measurement State

The action of a measurement on a quantum system can be captured through operators (see § 3.4.2). For a pure state $|\psi\rangle$ and a measurement operator M_j , the state after measurement is given by:

$$|\psi\rangle \rightarrow \frac{M_j |\psi\rangle}{\|M_j |\psi\rangle\|}. \quad (12.42)$$

For a mixed state represented by a density operator ρ , the transformation is as follows:

$$\rho \rightarrow \rho' = \frac{\sum_n p_n M_j |\psi_n\rangle \langle \psi_n| M_j^\dagger}{\text{tr} \left[\sum_n p_n M_j |\psi_n\rangle \langle \psi_n| M_j^\dagger \right]}. \quad (12.43)$$

Note that the normalization of a density operator is designed to produce a unity trace. This ensures that the probabilities of all possible outcomes of a measurement sum to 1. Simplifying using the procedure to prove Eq. 12.20, we obtain:

$$\rho \rightarrow \rho' = \frac{M_j \rho M_j^\dagger}{p_j}, \quad (12.44)$$

where p_j is the normalization factor or the probability of measuring M_j :

$$p_j = \text{tr}(M_j \rho M_j^\dagger). \quad (12.45)$$

Exercise 12.14 The state of a system, initially 50% in the state $|+\rangle$ and 50% in the state $|0\rangle$, is measured in the $\{|0\rangle, |1\rangle\}$ basis, and the outcome $|0\rangle$ is obtained. What is the new density operator of the system?

2 Ensemble of Measurement Outcomes

Instead of considering a single measurement operator M_j , we now extend our focus to a complete set of such operators $\{M_j\}$ (see § 3.4.2) that satisfy the completeness condition

$$\sum_j M_j^\dagger M_j = I. \quad (12.46)$$

The resulting ensemble of measurement outcomes is a mixed state represented by a density operator as follows:

$$\rho_M = \sum_j p_j \frac{M_j \rho M_j^\dagger}{p_j}. \quad (12.47)$$

Here, p_j is the probability of measuring M_j given by Eq. 12.45. Apparently, p_j cancels out, yielding the following crucial relationship:

$$\rho_M = \sum_j M_j \rho M_j^\dagger. \quad (12.48)$$

The ensemble of measurement outcomes encapsulates the probabilistic nature of quantum measurements. The density operator ρ_M represents the statistical mixture of all possible states that a quantum system can be in after being subjected to a measurement process. It offers a mechanism to represent the measurement and superposition principles of quantum mechanics statistically, thereby providing a complete picture of the system's possible states after a measurement.

The concept of the ensemble of measurement outcomes is closely related to the phenomenon of measurement-induced decoherence, a process that describes how quantum superpositions are reduced to classical mixtures as a result of measurement. A quantum system can exist in a superposition of states. When a measurement is performed, this superposition appears to collapse to a specific state corresponding to the measurement outcome. However, from a more comprehensive perspective, what happens is that the system becomes entangled with the measurement apparatus, leading to decoherence. The density operator ρ_M describes the state of the system after this decoherence has occurred.

Exercise 12.15 The state of a system, initially $|+\rangle$, is measured in the $\{|0\rangle, |1\rangle\}$ basis. What is the density operator of the post-measurement ensemble?

3 Measurements in an Orthonormal Basis

Now consider the special case where $\{M_j\}$ is the set of projection operators corresponding to an orthonormal basis $\{|\phi_j\rangle\}$:

$$M_j = |\phi_j\rangle\langle\phi_j|. \quad (12.49)$$

Applying the relationship from Eq. 12.48 for these projectors, we obtain the post-measurement mixed state as follows:

$$\rho_M = \sum_j |\phi_j\rangle\langle\phi_j| \rho |\phi_j\rangle\langle\phi_j| \quad (12.50a)$$

$$= \sum_j \langle\phi_j| \rho |\phi_j\rangle |\phi_j\rangle\langle\phi_j| \quad (12.50b)$$

$$= \sum_j p_j |\phi_j\rangle\langle\phi_j|, \quad (12.50c)$$

where

$$p_j = \langle\phi_j| \rho |\phi_j\rangle \quad (12.51)$$

is the probability of obtaining the outcome corresponding to $|\phi_j\rangle$.

State Collapse

The matrix ρ_M is diagonal in the basis $\{|\phi_j\rangle\}$, containing the probabilities p_j as its diagonal elements, while all off-diagonal elements are zero. If ρ was not initially diagonal in this basis, then ρ_M will differ from ρ . This difference is a consequence of the measurement process: a projective measurement collapses the system into a statistical mixture of the states $|\phi_j\rangle$ weighted by their respective probabilities p_j , effectively destroying any initial coherences between different basis states present in ρ . Hence, post-measurement, the system's state is a diagonal mixture of the measurement basis states. The system's final state coincides with the original ρ if and only if ρ was already diagonal in the chosen measurement basis.

4 POVM Measurements

Measurements in quantum mechanics can be more generally described using Positive Operator-Valued Measures (POVMs). A set of POVM operators $\{E_j\}$ satisfies two key conditions: each E_j is positive semidefinite ($E_j \geq 0$), and the sum of all POVM elements equals the identity operator ($\sum_j E_j = I$). These conditions guarantee that the calculated probabilities will sum to one. In this framework, each E_j corresponds to a distinct measurement outcome. The probability p_j of observing outcome j is given by:

$$p_j = \text{tr}(\rho E_j). \quad (12.52)$$

This probability can also be interpreted as the inner product $\langle \rho, E_j \rangle$ between the density operator ρ and the POVM element E_j .

In the traditional formulation of quantum measurements using a set of measurement operators $\{M_j\}$, each POVM element E_j is expressed as $E_j = M_j^\dagger M_j$. However, POVM measurements generalize this concept. The set of POVM operators $\{E_j\}$ can have a different cardinality from the dimension of the quantum state ρ , and the POVM elements are not required to be orthogonal. This broader framework is more aligned with the principles of quantum information theory.

12.2.7 * CPTP Maps and Quantum Channels

1 Completely Positive and Trace-Preserving (CPTP) Maps

Unitary evolution, as given by Eq. 12.33, and the quantum measurement process, described by Eq. 12.48, both transform the density operator of a quantum state. Since density operators must maintain a unit trace and be positive semidefinite, such transformations are referred to as Completely Positive and Trace-Preserving (CPTP) maps. In general, a CPTP map is an operation that transforms a density operator to another which can have a different dimensionality.

Definition

A CPTP map \mathcal{E} is a transformation that satisfies the following properties:

- Completely Positive: A map \mathcal{E} is said to be *positive* if for any density operator ρ , the map $\mathcal{E}(\rho)$ preserves the positive semidefiniteness of ρ , ensuring the physical validity of the transformed state.

A map \mathcal{E} is defined as *completely positive* if, when extended to any larger system that includes the original system as a subsystem, it preserves the positive

semidefiniteness of all states in this larger system. Formally, for any density operator ρ of the combined system (the original system plus an auxiliary system), the extended map $(\mathcal{E} \otimes I)(\rho)$ is positive, where I is the identity map on the auxiliary system.

- **Trace-Preserving:** For any density operator ρ , $\text{tr}(\mathcal{E}(\rho)) = \text{tr}(\rho)$. This property ensures that the total probability (sum of probabilities of all outcomes) is conserved in the transformation.
- **Linearity:** $\mathcal{E}(a\rho + b\sigma) = a\mathcal{E}(\rho) + b\mathcal{E}(\sigma)$ for any density operators ρ, σ , and scalars a, b .

Applications

CPTP maps offer a comprehensive framework for describing the dynamics of quantum states, and are a fundamental tool in quantum computation and information theory. They encompass a wide range of physical processes:

- **Unitary Evolutions:** Unitary operations, such as the action of quantum gates, are reversible and a special case of CPTP maps.
- **Quantum Measurements:** These maps model the process of quantum measurement, including the decoherence effect, as in Eq. 12.48.
- **Simulation of Open Systems.** CPTP maps can model open quantum systems where interactions with external environments are significant. They facilitate the inclusion of auxiliary systems in the analysis, which serve to represent environmental effects. In this context, an auxiliary system functions as a supplementary quantum system, considered in conjunction with the primary system to simulate the combined effect of the environment and the system dynamics.
- **Reduction of Systems via Partial Trace:** CPTP maps can represent the process of tracing out part of a quantum system, effectively reducing it to a subsystem. The partial trace operation ensures that the reduced state remains a valid density operator.
- **Other Quantum Operations:** These maps include a broad class of quantum operations, such as those involving interactions with an environment. These operations can lead to various effects like quantum noise, dissipation, and more general quantum dynamics beyond purely unitary transformations or simple measurements.

2 Quantum Channels

CPTP maps can be modeled using quantum channels, which are often represented by Kraus decompositions or operator-sum representations. A Kraus decomposition provides an intuitive and computationally convenient method to express a CPTP map by breaking it down into a set of operations on the quantum state. Thus, a quantum channel, describing the evolution of a quantum state under a CPTP map, is represented as:

$$\mathcal{E}(\rho) = \sum_j s_j K_j \rho K_j^\dagger, \quad (12.53)$$

where the K_j are Kraus operators that describe the quantum channel, and s_j is the probability that the transformation represented by K_j occurs. Here K_j does not have to be a square matrix.

These operators must satisfy the modified completeness relation:

$$\sum_j s_j K_j^\dagger K_j = I, \quad (12.54)$$

ensuring that the map \mathcal{E} is trace-preserving.

An alternative representation, which has the probability s_j factored into the operator K_j : $\sqrt{s_j} K_j \mapsto K_j$, and thus does not explicitly use the probabilities s_j , is given by:

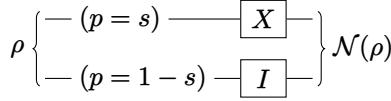
$$\mathcal{E}(\rho) = \sum_j K_j \rho K_j^\dagger, \quad (12.55)$$

with the completeness condition:

$$\sum_j K_j^\dagger K_j = I. \quad (12.56)$$

We have seen that quantum measurement can be represented as a Kraus decomposition in Eq. 12.48. Below are additional examples:

■ **Example 12.6 — Bit-Flip Noise Channel.** A bit-flip error flips the state of a qubit from $|0\rangle$ to $|1\rangle$, or vice versa, with some specified probability, s . It can be modeled by applying an X gate, randomly with probability s , as illustrated below:



The Kraus representation for the associated noise channel is:

$$\mathcal{N}(\rho) = (1-s)I\rho I + sX\rho X = (1-s)\rho + sX\rho X. \quad (12.57)$$

We will delve into such random noise channels in the next section. ■

■ **Example 12.7 — Complete Dephasing Noise Channel.** Consider the general qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with density operator

$$\rho = |\psi\rangle\langle\psi| = \begin{bmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{bmatrix}. \quad (12.58)$$

If we measure it in the standard basis, but actually “forget” the outcome, we can see that the resulting state will output $|0\rangle$ with output probability $|\alpha|^2$, and output $|1\rangle$ with probability $|\beta|^2$. Thus, the output is:

$$\mathcal{N}(\rho) = \begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}. \quad (12.59)$$

The effect is to zero out all the off-diagonal terms in ρ , thus destroying all coherence. This process is often referred to as complete dephasing. ■

Exercise 12.16 Consider an initial state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. If a bit flip is introduced to the system with a random probability of occurrence s , determine the resulting density operator after this noise channel acts on the state.

3 Stinespring Dilation: Unitary Embedding of a CPTP Map

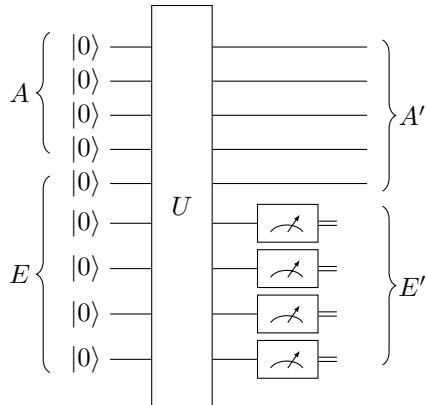


Figure 12.3: Stinespring Dilation: Unitary Embedding of a CPTP Map

Stinespring dilation provides a way to visualize and understand general quantum operations (CPTP maps) as unitary evolutions in a larger space, which includes the system and the environment. Stinespring dilation and Kraus decomposition are mathematically equivalent, yet they provide distinct perspectives on the nature of quantum operations.

Stinespring dilation can be illustrated using the quantum circuit in Fig. 12.3. In this schematic, qubits in group A are considered the input qubits of the system, while those in group E function as ancillary qubits, simulating an environment. The combined system $A + E$ is subject to a unitary transformation U . Following this transformation, the qubits are categorized into groups A' and E' , which correspond to the transformed states of A and E , respectively. The qubits in E' are subsequently measured (typically in the computational basis) and then discarded (traced out), while the remaining qubits in A' constitute the output of the CPTP map.

This process is encapsulated by the Stinespring Dilation Theorem, which asserts that any CPTP map can be implemented as a unitary operation on an expanded Hilbert space that includes an ancillary system, followed by tracing out some degrees of freedom—specifically, those associated with the environment. This theorem demonstrates that operations which appear non-unitary, such as decoherence or measurement, can be modeled within the framework of unitary quantum mechanics by accounting for interactions with an ancillary system or external environment.

Formally, the Stinespring Dilation Theorem states that given a CPTP map \mathcal{E} acting on a state ρ in system A , there exists an ancillary system E , a unitary operation U , and a state $|\psi\rangle$ in the combined system such that:

$$\mathcal{E}(\rho) = \text{tr}_E (U(\rho \otimes |\psi\rangle\langle\psi|)U^\dagger), \quad (12.60)$$

where tr_E denotes the partial trace (discussed in § 12.2.8) over the ancillary system E , effectively capturing the essence of the CPTP map's action on the state ρ .

A common question arises: If unitary operations are inherently reversible, does that imply CPTP maps, including processes like decoherence, are also reversible? The answer is “No.” The key lies in the irreversible act of tracing out the environment subsystem E' after the unitary operation. This step, essential for modeling the effect of the environment, discards information, making the process inherently irreversible. Thus, despite the unitary transformation's reversibility in the combined system $A + E$, the practical application of CPTP maps remains irreversible due to the loss of access to the environment's state post-interaction, highlighting a fundamental quantum-to-classical transition.

12.2.8 * Partial Trace and Reduced Density Operators

The concept of the partial trace of density operators in quantum mechanics bears a strong analogy to marginal probability in classical probability theory (see § 13.1.4). Both methodologies enable focusing on a subset within a larger system by summing over and thereby effectively disregarding the other components of the system.

Partial trace serves a complementary role to tensor product. While tensor product is used to construct a composite quantum system from individual states, partial trace allows us to extract and examine the state of a specific subsystem, effectively removing information about other parts of the system.

Partial trace is a fundamental tool for describing parts of larger quantum systems. Its applications range widely, including understanding quantum entanglement, studying the effects of environmental interactions on quantum systems (such as decoherence), and analyzing quantum information processes.

1 Definition

Consider a bipartite composite system represented by a density operator ρ_{AB} . The partial trace of ρ_{AB} over subsystem B is defined as:

$$\rho_A \equiv \text{tr}_B(\rho_{AB}) = \sum_i (I_A \otimes \langle i|_B) \rho_{AB} (I_A \otimes |i\rangle_B), \quad (12.61)$$

or, in the partial product notation (see Appendix C),

$$\text{tr}_B(\rho_{AB}) = \sum_i \langle i|_B \rho_{AB} |i\rangle_B, \quad (12.62)$$

where $\{|i\rangle_B\}$ is any orthonormal basis for the Hilbert space of subsystem B . In this case, the result of the partial trace, ρ_A , is also referred to as the reduced density operator of subsystem A .

The reduced density operator ρ_A represents the state of subsystem A after averaging over the degrees of freedom of subsystem B . This operation effectively encompasses all possible outcomes of partial measurements on subsystem B , providing a statistical description of subsystem A irrespective of the specific state or measurement outcome of subsystem B .

Similarly, the partial trace over subsystem A , or the reduced density operator of subsystem B , is defined as:

$$\rho_B \equiv \text{tr}_A(\rho_{AB}) = \sum_i \langle i|_A \rho_{AB} |i\rangle_A, \quad (12.63)$$

where $\{|i\rangle_A\}$ is any orthonormal basis for the Hilbert space of subsystem A .

Exercise 12.17 The reduced density operator given in Eq. 12.62 remains the same regardless of the chosen orthonormal basis for the Hilbert space of subsystem B . Demonstrate this.

2 Independent Subsystems

In general, $\rho_{AB} \neq \rho_A \otimes \rho_B$. However, if the subsystems are independent, then $\rho_{AB} = \rho_A \otimes \rho_B$, which can be demonstrated as follows. In this special case, Eq. 12.61 simplifies to

$$\text{tr}_B(\rho_{AB}) = \sum_i (I_A \otimes \langle i|_B)(\rho_A \otimes \rho_B)(I_A \otimes |i\rangle_B) \quad (12.64a)$$

$$= \rho_A \otimes \sum_i \langle i|_B \rho_B |i\rangle_B \quad (12.64b)$$

$$= \rho_A \otimes \text{tr}(\rho_B) \quad (12.64c)$$

$$= \rho_A, \quad (12.64d)$$

as expected, since $\text{tr}(\rho_B) = 1$.

3 Examples

In the following, we will explore partial trace through several examples. We begin with a visual representation of matrix elements in a two-qubit system to establish a foundational understanding. Subsequently, we examine the partial trace in different contexts: a correlated mixed state to demonstrate how classical correlations are handled, and a Bell state to reveal the intriguing nature of quantum entanglement. We then extend our exploration to consider the effects of local operations in entangled systems, which leads us to a deeper understanding of quantum nonlocality.

■ Example 12.8 — General Two-Qubit System: Matrix Visualization. The density operator of a generic two-qubit system can be represented as

$$\rho_{AB} = \begin{bmatrix} \langle 00| & \langle 01| & \langle 10| & \langle 11| \\ |00\rangle & \rho_{0000} & \rho_{0001} & \rho_{0010} & \rho_{0011} \\ |01\rangle & \rho_{0100} & \rho_{0101} & \rho_{0110} & \rho_{0111} \\ |10\rangle & \rho_{1000} & \rho_{1001} & \rho_{1010} & \rho_{1011} \\ |11\rangle & \rho_{1100} & \rho_{1101} & \rho_{1110} & \rho_{1111} \end{bmatrix}. \quad (12.65)$$

Here, for clearer visualization, we have used binary indices 00, 01, 10, 11 (instead of 1, 2, 3, 4) to label the matrix elements.

To compute ρ_A , the reduced density operator for qubit A , we perform the partial trace of ρ_{AB} over subsystem B . This process can be visualized as combining elements of ρ_{AB} that correspond to the same state of qubit A . For each state of qubit A , we add together the elements of ρ_{AB} where qubit B is in either state $|0\rangle$ or $|1\rangle$:

$$\rho_A = \text{tr}_B(\rho_{AB}) = \begin{bmatrix} \rho_{0000} + \rho_{0101} & \rho_{0010} + \rho_{0111} \\ \rho_{1000} + \rho_{1101} & \rho_{1010} + \rho_{1111} \end{bmatrix}. \quad (12.66)$$

Here, the element $\rho_{0000} + \rho_{0101}$, for example, represents the sum of probabilities where qubit A is in state $|0\rangle$, irrespective of the state of qubit B . It corresponds to $\langle 0|_B (\rho_{0000} |00\rangle\langle 00|) |0\rangle_B + \langle 1|_B (\rho_{0101} |01\rangle\langle 01|) |1\rangle_B$ in Eq. 12.62. This ‘partial tracing’ effectively removes the detailed information about qubit B , leaving us with the reduced state of qubit A .

Similarly, the reduced density operator for qubit B , ρ_B , can be obtained by summing the matrix elements corresponding to the states of qubit A :

$$\rho_B = \text{tr}_A(\rho_{AB}) = \begin{bmatrix} \rho_{0000} + \rho_{1010} & \rho_{0001} + \rho_{1011} \\ \rho_{0100} + \rho_{1110} & \rho_{0101} + \rho_{1111} \end{bmatrix}. \quad (12.67)$$

■

Exercise 12.18 Consider a two-qubit product state $|\psi\rangle = |++\rangle$, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.

- (a) Express $|\psi\rangle$ in terms of $|0\rangle$ and $|1\rangle$ and compute the density operator $\rho_{AB} = |++\rangle\langle ++|$.
- (b) Compute the reduced density operators ρ_A and ρ_B from the partial trace of ρ_{AB} . Verify that $\rho_A = \rho_B = |+\rangle\langle +|$.
- (c) Verify that $\rho_{AB} = \rho_A \otimes \rho_B$.

This exercise is designed to show that the reduced density operators derived from partial tracing align with our expectations for simple product states.

■ **Example 12.9 — Partial Trace of a Correlated Mixed State.** Consider a two-qubit system with the density operator ρ_{AB} given by:

$$\rho_{AB} = \frac{1}{2} |00\rangle\langle 00| + \frac{1}{2} |11\rangle\langle 11|, \quad (12.68)$$

representing a mixed state with classical correlations between the two qubits. It is not an entangled state.

We want to find the reduced density operator for qubit A . The partial trace over qubit B is calculated as:

$$\text{tr}_B(\rho_{AB}) = \langle 0|_B \rho_{AB} |0\rangle_B + \langle 1|_B \rho_{AB} |1\rangle_B. \quad (12.69)$$

Substituting the value of ρ_{AB} , we get:

$$\text{tr}_B(\rho_{AB}) = \langle 0|_B \left(\frac{1}{2} |00\rangle\langle 00| + \frac{1}{2} |11\rangle\langle 11| \right) |0\rangle_B \quad (12.70a)$$

$$+ \langle 1|_B \left(\frac{1}{2} |00\rangle\langle 00| + \frac{1}{2} |11\rangle\langle 11| \right) |1\rangle_B$$

$$= \frac{1}{2} \langle 0|_B |00\rangle\langle 00| |0\rangle_B + \frac{1}{2} \langle 1|_B |11\rangle\langle 11| |1\rangle_B \quad (12.70b)$$

$$= \frac{1}{2} |0\rangle\langle 0| + \frac{1}{2} |1\rangle\langle 1| \quad (12.70c)$$

$$= \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (12.70d)$$

Therefore, the reduced density operator for qubit A is:

$$\rho_A = \frac{I}{2}, \quad (12.71)$$

which indicates that qubit A is in a maximally mixed state. Similarly, $\rho_B = \frac{I}{2}$. ■

■ **Example 12.10 — Partial Trace of a Bell State.** The Bell state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ has a density operator given by:

$$\rho_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)(\langle 00| + \langle 11|). \quad (12.72)$$

Expanding the terms yields:

$$\rho_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|). \quad (12.73)$$

Let's now calculate the partial trace over the B subsystem, using the computational basis as our orthonormal basis $\{|i\rangle\}$:

$$\rho_A = \text{tr}_B(\rho_{AB}) \quad (12.74a)$$

$$= \frac{1}{2} (\langle 0|_B \rho_{AB} |0\rangle_B + \langle 1|_B \rho_{AB} |1\rangle_B) \quad (12.74b)$$

$$= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \quad (12.74c)$$

$$= \frac{I}{2}. \quad (12.74d)$$

Similarly, $\rho_B = \frac{I}{2}$.

Apparently, both ρ_A and ρ_B represent a uniform mixed state with 50% probability of $|0\rangle$ and 50% probability of $|1\rangle$, a surprising result considering that the joint state of A and B is a pure, entangled state. This illuminates the non-intuitive properties of quantum entanglement, where subsystems can be mixed even when the global system is in a pure state.

Comparing with Example 12.9, we see that the reduced state of a subsystem can be maximally mixed, regardless of whether the overall system is in a correlated mixed state or an entangled pure state. Thus, in quantum mechanics, the properties of parts of a system do not always straightforwardly reflect the properties of the whole system.

Can we differentiate between this Bell state and the classically correlated mixed state in Example 12.9? The distinction becomes clear with the application of the Bell inequality and the Bell test experiments. For a more comprehensive understanding of these concepts, refer to Chapter 9. ■

Exercise 12.19 Compute ρ_A and ρ_B for the Bell state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, as in Example 12.10, but now using the $\{|+\rangle, |-\rangle\}$ basis, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.

■ **Example 12.11 — Local Operations in Entangled Systems.** Let's continue from the previous example, and consider the scenario where Alice and Bob are spatially separated and share a Bell pair described by the density operator in Eq. 12.72. Suppose Bob performs a local operation on his qubit, such as a measurement or a state change. We analyze whether Alice can perceive this action through her part of the Bell pair using the concept of the partial trace.

Case 1: Local Unitary Operations

Suppose Bob performs a local unitary operation U_B on his qubit. The state of the system after this operation is given by:

$$\rho'_{AB} = (I_A \otimes U_B)\rho_{AB}(I_A \otimes U_B^\dagger). \quad (12.75)$$

Alice's reduced state, obtained by tracing out Bob's subsystem, is $\rho'_A = \text{tr}_B(\rho'_{AB})$. Since the trace operation is invariant under unitary transformations, $\text{tr}_B(\rho'_{AB}) = \text{tr}_B(\rho_{AB})$. Therefore,

$$\rho'_A = \rho_A = \frac{I}{2}, \quad (12.76)$$

indicating that Alice's reduced state is unaffected. Although the unitary operation changes the state of Bob's qubit, it does not collapse the quantum state, nor does it change the entanglement properties in a way observable to Alice.

Case 2: Local Measurements

Now, consider Bob performs a local measurement. This collapses the state of the system into one of the eigenstates corresponding to the measurement outcome. However, without further communication, Alice would not know which eigenstate Bob's qubit has collapsed to. So she would experience a mixed state corresponding to the statistical ensemble of both measurement outcomes, given by the density operator in Eq. 12.68. As we have worked out in Example 12.9, Alice's reduced density operator is $\frac{I}{2}$.

Thus, Alice's observable statistics remain consistent with the reduced density operator calculated before Bob's actions. She cannot distinguish whether her mixed state is due to the original entanglement or the collapse caused by Bob's measurement, highlighting the non-observability of Bob's local measurements from Alice's perspective without additional classical communication.

This example illustrates a fundamental aspect of quantum mechanics: local operations on part of an entangled system do not have observable consequences on the distant part of the system, and entanglement itself does not allow for superluminal

communication. This underscores the principles of the no-communication theorem in quantum mechanics. ■

■ **Example 12.12 — GHZ State and the Monogamy of Entanglement.** Consider the GHZ state $|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$, which extends the concept of entanglement to three qubits, held by Alice, Bob, and Charlie respectively. Unlike pairs of qubits in the Bell state, the three qubits in the GHZ state are mutually entangled; measuring one qubit instantly determines the states of the other two. Its density operator is given by:

$$\rho_{ABC} = \frac{1}{2}(|000\rangle + |111\rangle)(\langle 000| + \langle 111|). \quad (12.77)$$

What happens if we observe just Alice and Bob's qubits without Charlie's? To find out, we perform a partial trace over qubit C , yielding:

$$\rho_{AB} = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|. \quad (12.78)$$

Surprisingly, ρ_{AB} resembles the density operator of a classically correlated mixed state, not the entangled Bell state one might expect. This shows that the entanglement in the three-qubit GHZ state is fully manifested only when all three qubits are considered together; with only two, we observe merely classical correlations.

This phenomenon is often referred to as the monogamy of entanglement, highlighting its “all-or-nothing” nature. As a result, the triple cannot share additional entanglement with any external system. Attempting to entangle a fourth qubit with the GHZ state would inevitably weaken its intrinsic entanglement. The principle of monogamy underlies the security of certain quantum cryptography protocols, where any eavesdropping attempt on part of an entangled state disrupts the entanglement in a detectable manner.

Another interesting three-qubit entangled state is the W state, $|\text{W}\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$. In the W state, there is partial entanglement between each pair of qubits, but none are maximally entangled. Readers are encouraged to explore this further by performing a partial trace on the density operator of the W state. ■

12.2.9 *Extension of No-Cloning and No-Communication Theorems

The no-cloning and no-communication theorems, initially introduced within the context of unitary operations in closed quantum systems, are equally applicable to more general quantum operations defined by Completely Positive Trace-Preserving (CPTP) operations. These operations encompass measurements and other non-reversible effects, thereby broadening the applicability of these fundamental theorems. In this subsection, we outline proofs of the no-cloning and no-communication theorems within the framework of CPTP operations.

1 The No-Cloning Theorem: A Proof Outline

The No-Cloning Theorem (see § 5.1.3) asserts that it is impossible to create an identical copy of an arbitrary unknown quantum state. A proof was provided under unitary transformations in § 5.7. The proof can be extended to CPTP maps, which can include measurements and other quantum operations.

Consider an arbitrary quantum state, $|\psi\rangle$. Assume there exists a CPTP map \mathcal{E} that can clone any quantum state, such that $\mathcal{E}(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|$. The map \mathcal{E} must simultaneously clone $|0\rangle$, $|1\rangle$, and $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$:

$$\mathcal{E}(|0\rangle\langle 0|) = |0\rangle\langle 0| \otimes |0\rangle\langle 0|, \quad (12.79a)$$

$$\mathcal{E}(|1\rangle\langle 1|) = |1\rangle\langle 1| \otimes |1\rangle\langle 1|, \quad (12.79b)$$

$$\mathcal{E}(|+\rangle\langle +|) = |+\rangle\langle +| \otimes |+\rangle\langle +|. \quad (12.79c)$$

However, if we add the first two equations and multiply the result by $\frac{1}{2}$, we do not get the same result as the third equation. This discrepancy highlights the contradiction and proves that a universal cloning machine cannot exist within the framework of CPTP maps, thus proving the no-cloning theorem.

2 The No-Communication Theorem: A Proof Outline

The No-Communication Theorem (see § 9.7) states that during the measurement of an entangled quantum state, it is fundamentally impossible for one observer to communicate information to another solely through the act of measurement. Here is an outline of the proof of this Theorem, building on the previous example.

Suppose we have a composite quantum system described by a joint density operator ρ_{AB} . Let \mathcal{E} be a completely positive, trace-preserving (CPTP) map that acts only on subsystem B . Physically, \mathcal{E} can represent a quantum operation (such as a measurement or a unitary operation) performed by Bob on his subsystem.

Now, consider the reduced density operator of subsystem A . Before Bob's operation, it is given by $\text{tr}_B(\rho_{AB})$. After Bob's operation, the state of the total system becomes:

$$\rho'_{AB} = (I \otimes \mathcal{E})(\rho_{AB}), \quad (12.80)$$

where I is the identity map acting on subsystem A .

The reduced density operator of subsystem A is then

$$\text{tr}_B(\rho'_{AB}) = \text{tr}_B[(I \otimes \mathcal{E})(\rho_{AB})]. \quad (12.81)$$

A CPTP map \mathcal{E} maintains the trace of a density operator, i.e., $\text{tr}(\mathcal{E}\rho) = \text{tr}(\rho)$. When taking the partial trace over B , we are effectively summing over the degrees of freedom of subsystem B . Since \mathcal{E} is trace-preserving and only acts on B , the overall trace of the entire system over B remains unchanged. Thus, we have

$$\text{tr}_B[(I \otimes \mathcal{E})(\rho_{AB})] = \text{tr}_B(\rho_{AB}). \quad (12.82)$$

Therefore,

$$\text{tr}_B(\rho'_{AB}) = \text{tr}_B(\rho_{AB}). \quad (12.83)$$

This result indicates that the reduced density operator of subsystem A (i.e., Alice's subsystem) remains unchanged regardless of what operation Bob performs on subsystem B . Consequently, Bob cannot send any information to Alice through his actions on his subsystem.

This proof assumes that Alice and Bob's subsystems do not interact during the process. If there is an interaction, information can be transferred, but it cannot occur faster than light as required by the theory of relativity.

12.3 Error Mechanisms in Quantum Computing

In quantum computing, especially in the Noisy Intermediate-Scale Quantum (NISQ) era, errors can be broadly divided into coherent and incoherent errors, which are the focus of the section. These errors manifest differently in the dynamics of the quantum system and, consequently, have different implications for error correction and fault tolerance. This section also introduces merit parameters that serve as quantitative metrics for characterizing the impact and prevalence of these errors.

12.3.1 *Coherent Errors

Coherent errors typically stem from inaccuracies in control parameters such as pulse amplitude, duration, and frequency. They often manifest as unitary transformations that deviate from the ideal transformations. Below, we discuss two main types of coherent errors: gate calibration errors and cross-talk errors.

1 Gate Calibration Errors

Gate calibration errors arise when there is a discrepancy between the ideal and the actual implementations of a quantum gate. These errors are frequently due to inaccuracies in control parameters such as pulse amplitude, frequency, or timing.

To illustrate, consider a simple X -gate applied to a qubit initially in the state $|0\rangle$. In an ideal situation, the gate would transform the qubit into the state $|1\rangle$. Repeated applications of an ideal X -gate would result in the qubit state oscillating between $|0\rangle$ and $|1\rangle$, as shown in Fig. 12.4.

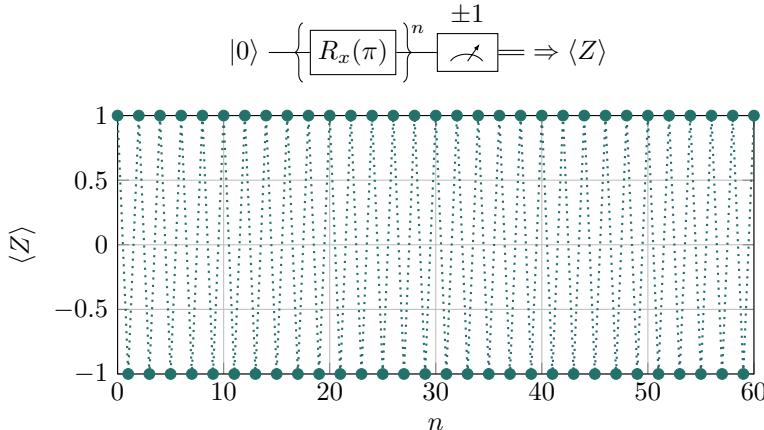


Figure 12.4: Repeated Application of an Ideal X Gate

The X -gate is typically implemented as a π rotation around the x -axis: $X = iR_x(\pi)$ (see §§ 4.5.2 and 5.2.5). However, calibration errors can occur. For instance, the amplitude of the control pulse may be slightly higher than required, resulting

in an actual operation $R_x(\pi + \varepsilon)$. When we repeat this off-calibration X -gate, the resulting qubit state will not simply oscillate between $|0\rangle$ and $|1\rangle$, but may evolve into a superposition state. This is illustrated in Fig. 12.5.

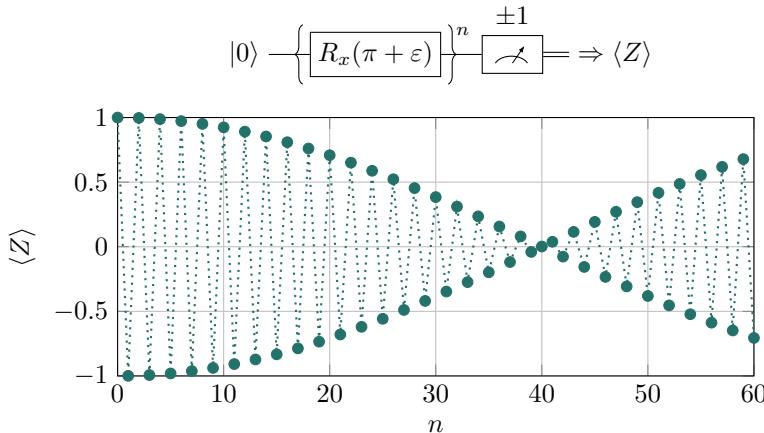


Figure 12.5: Repeated Application of an Off-Calibrated X Gate

Exercise 12.20 Deduce the value of the calibration error ε based on the graph in Fig. 12.5. Hint: $(R_x(\pi + \varepsilon))^n = R_x(n\pi + n\varepsilon)$. When $n\varepsilon$ accumulates to $\frac{\pi}{2}$, the final state is a uniform superposition, yielding $\langle Z \rangle = 0$.

As demonstrated in the preceding example, the impact of gate calibration errors can be exacerbated in circuits with increasing depth (n). Specifically, as n grows, these errors may accumulate, leading to significant deviations in the final computational results from the expected outcomes.

2 Cross-Talk Errors

Cross-talk errors occur due to unintended interactions between qubits that are either adjacent or otherwise coupled. These interactions can introduce undesired rotations and entanglement, thereby affecting the fidelity of gate operations.

Consider, for instance, an unwanted ZZ coupling governed by the following two-qubit Hamiltonian:

$$H_{ZZ} = -\frac{\hbar\omega}{2} ZZ, \quad (12.84)$$

where ω represents the coupling strength, and ZZ denotes $Z \otimes Z$.

This Hamiltonian leads to a unitary evolution given by $R_{zz}(\omega t)$, which can be viewed as an unwanted $ZZ(\theta)$ gate with $\theta = \omega t$:

$$R_{zz}(\omega t) = \cos \frac{\omega t}{2} I + i \sin \frac{\omega t}{2} ZZ. \quad (12.85)$$

Such an interaction results in a phase walk-off in the computational basis. To measure this effect, we sandwich the $R_{zz}(\omega t)$ unitary between Hadamard gates, as

depicted in Fig. 12.6. In the absence of coupling ($\omega = 0$), the measurement on the first qubit yields $\langle ZI \rangle = 1$, corresponding to the expected outcome $|00\rangle$.

However, with coupling, the measurement outcome changes to $\langle ZI \rangle = \cos \omega t$, indicating the emergence of unwanted phase shift that deteriorates over time. (Similarly for $\langle IZ \rangle$ measured on the second qubit.)

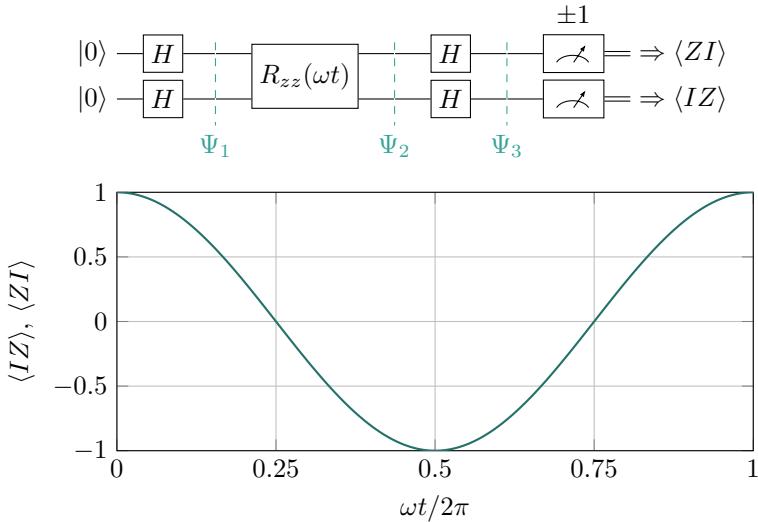


Figure 12.6: Effect of ZZ Cross-Talk over Time

Below is an outline for the derivation of $\langle ZI \rangle = \cos \omega t$:

$$|\Psi_1\rangle = HH|00\rangle = |+\rangle, \quad (12.86a)$$

$$|\Psi_2\rangle = R_{zz}(\omega t)|\Psi_1\rangle = \cos \frac{\omega t}{2}|+\rangle + i \sin \frac{\omega t}{2}|-\rangle, \quad (12.86b)$$

$$|\Psi_3\rangle = HH|\Psi_2\rangle = \cos \frac{\omega t}{2}|00\rangle + i \sin \frac{\omega t}{2}|11\rangle, \quad (12.86c)$$

where we have used the following properties: $H|0\rangle = |+\rangle$, $H|1\rangle = |-\rangle$, $H|+\rangle = |0\rangle$, $H|-\rangle = |1\rangle$, $Z|+\rangle = |-\rangle$, $Z|-\rangle = |+\rangle$.

$$\langle ZI \rangle = \langle \Psi_3 | ZI | \Psi_3 \rangle \quad (12.87a)$$

$$= \cos^2 \frac{\omega t}{2} - \sin^2 \frac{\omega t}{2} = \cos \omega t. \quad (12.87b)$$

Exercise 12.21 Plot $\langle ZZ \rangle$ as function of ωt .

12.3.2 *Incoherent Errors

Incoherent errors are fundamentally stochastic in nature and primarily arise due to uncontrolled interactions between the quantum system and its external environment. Unlike coherent errors, which can be modeled deterministically, incoherent errors are statistically represented using density operators and noise channels.

There are numerous types of incoherent errors that affect both single-qubit and multi-qubit systems. In the following subsections, we will focus on three key examples: bit-flip, phase-flip, and depolarizing errors. We will also outline the noise channel representations for other types of errors. The concepts and formulas presented here are essential for understanding and simulating noise in quantum computing.

Pure States, Mixed States, Noise, and Noise Channels

Here is a recapitulation of some key concepts from § 12.2.

Pure States: In quantum mechanics, a pure state is described by a unique ket vector $|\psi\rangle$ in a Hilbert space. For a pure state, the associated density operator ρ is a projection operator given by $\rho = |\psi\rangle\langle\psi|$. A pure state is free from ‘quantum noise,’ meaning that repeated quantum measurements, under identical conditions, yield the same probabilities for each possible outcome, which in the case of certain observables can produce deterministic results in quantum computing.

Mixed States: In contrast, a mixed state is a statistical ensemble of multiple potential quantum states. It is represented by a density operator ρ that is not expressible as $|\psi\rangle\langle\psi|$ for any ket $|\psi\rangle$. A mixed state introduces classical uncertainty, leading to non-deterministic results in quantum computing because the system can be in any of the states of the ensemble with a certain probability.

Noise and Errors: In quantum systems, “noise” often refers to processes that convert pure states into mixed states. Such processes include decoherence, dissipation, and interaction with the environment, among others. This random noise manifests as computational errors, which can affect the outcomes of quantum computations.

Noise Channels and Kraus Representation: Noise and errors in quantum computing are commonly modeled through quantum channels, which act as trace-preserving transformations on density operators. This is also known as the Kraus representation:

$$\mathcal{N}(\rho) = \sum_j s_j K_j \rho K_j^\dagger,$$

where $\{K_j\}$ are the Kraus operators and $\{s_j\}$ are the corresponding probabilities.

This representation serves as a powerful framework for both understanding and mitigating the impacts of noise in quantum systems.

1 Bit-Flip Error

A bit-flip error inverts the state of a qubit from $|0\rangle$ to $|1\rangle$ or vice versa with a predetermined probability s . This error can be modeled by applying an X gate randomly with probability s . The corresponding noise channel for a bit-flip error is represented as

$$\mathcal{N}_{\text{bit-flip}}(\rho) = (1 - s)\rho + sX\rho X. \quad (12.88)$$

As discussed in Example 12.5, an X gate transforms a general quantum state, represented by the Bloch vector (x, y, z) , into $(x, -y, -z)$. By combining both terms in $\mathcal{N}_{\text{bit-flip}}(\rho)$, we obtain:

$$\mathcal{N}_{\text{bit-flip}}(x, y, z) = (x, (1 - 2s)y, (1 - 2s)z). \quad (12.89)$$

Hence, the influence of bit-flip errors can be visualized as a contraction of the terminal points of the qubit states in the Bloch sphere along the y and z axes by a factor of $1 - 2s$, as depicted in Fig. 12.7.

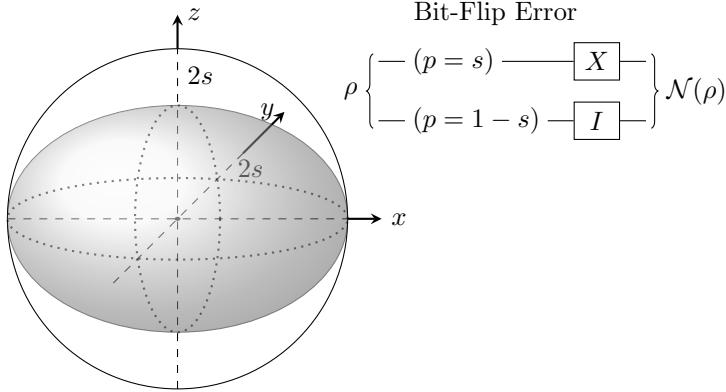


Figure 12.7: Bit-Flip Error in the Bloch Sphere

To elucidate the impact of bit-flip errors further, consider their combined effect with a sequence of X (or $R_x(\pi)$) gates, as presented in Fig. 12.4. The resulting Kraus representation becomes:

$$\mathcal{N}_{X\text{-and-BF}}(\rho) = X\mathcal{N}_{\text{bit-flip}}(\rho)X = (1 - s)X\rho X + s\rho, \quad (12.90)$$

or, when expressed in terms of Bloch vectors,

$$\mathcal{N}_{X\text{-and-BF}}(x, y, z) = (x, (2s - 1)y, (2s - 1)z). \quad (12.91)$$

Upon n repeated applications of this transformation, one obtains:

$$\mathcal{N}_{X\text{-and-BF}}^n(x, y, z) = (x, (2s - 1)^n y, (2s - 1)^n z). \quad (12.92)$$

Assuming the initial state is $|0\rangle$, corresponding to $(0, 0, 1)$, we have:

$$\mathcal{N}_{X\text{-and-BF}}^n(0, 0, 1) = (0, 0, (2s - 1)^n). \quad (12.93)$$

Therefore,

$$\langle Z \rangle = (2s - 1)^n = (-1)^n(1 - 2s)^n. \quad (12.94)$$

The above equation is graphed in Fig. 12.8 with $s = 1\%$. While the unaltered application of the X gate results in an oscillation between the qubit states $|0\rangle$ and $|1\rangle$ (see Fig. 12.4), the presence of bit-flip errors induces a damping effect on the amplitude of this oscillation by the factor $(1 - 2s)^n$.

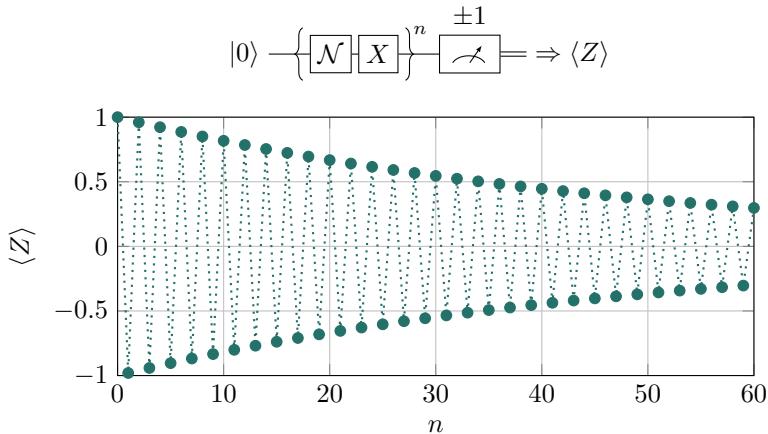


Figure 12.8: Repeated Applications of an X Gate with Bit-Flip Noise

2 Phase-Flip Error

A phase-flip error results in a sign change in the $|1\rangle$ state of a qubit, and is represented by the random application of the Pauli-Z operator with a predetermined probability s . The Kraus representation of the phase-flip error is given by:

$$\mathcal{N}_{\text{phase-flip}}(\rho) = (1 - s)\rho + sZ\rho Z. \quad (12.95)$$

The Z gate modifies a general quantum state represented by the Bloch vector (x, y, z) into $(-x, -y, z)$. Integrating both terms in $\mathcal{N}_{\text{phase-flip}}(\rho)$, we find:

$$\mathcal{N}_{\text{phase-flip}}(x, y, z) = ((1 - 2s)x, (1 - 2s)y, z). \quad (12.96)$$

Thus, phase-flip errors can be visualized as a contraction of the terminal points of the qubit states along the x and y axes in the Bloch sphere by a factor of $1 - 2s$, as depicted in Fig. 12.9.

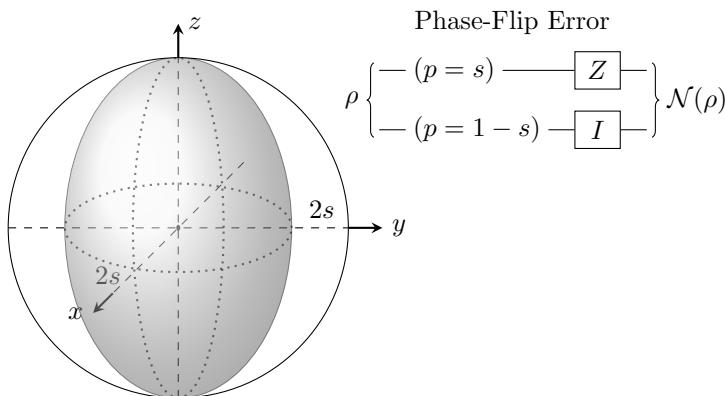


Figure 12.9: Phase-Flip Error in the Bloch Sphere

3 Depolarizing Error

Depolarizing error describes the transition of a qubit from a pure state into a completely mixed state comprising $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$. In this process, the qubit loses both its phase information and the superposition between its basis states. The Kraus representation of depolarizing error is:

$$\mathcal{N}_{\text{dephase}}(\rho) = (1 - s)\rho + \frac{s}{3}(X\rho X + Y\rho Y + Z\rho Z), \quad (12.97)$$

where the parameter s quantifies the noise strength.

This error manifests as a uniform contraction of the terminal points of the qubit states within the Bloch sphere, as illustrated in Fig. 12.10.

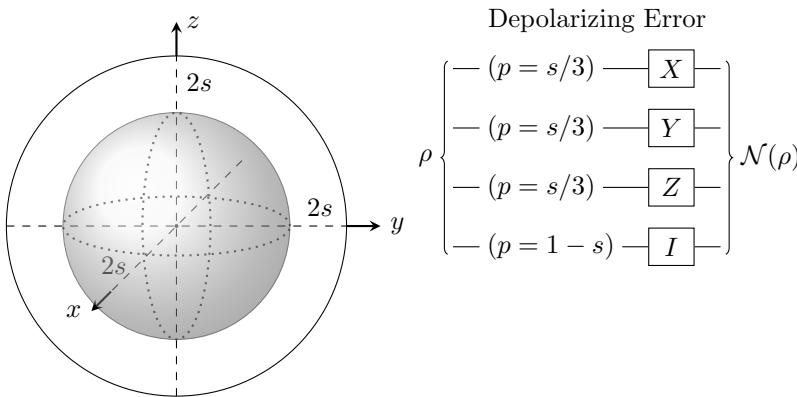


Figure 12.10: Depolarizing Error in the Bloch Sphere

4 Dephasing Error

Dephasing error occurs when the relative phase between the basis states $|0\rangle$ and $|1\rangle$ becomes randomized. The phase-flip error can be considered a special case of dephasing error. Dephasing error can be modeled mathematically by applying a random phase using the $R_z(\theta)$ gate to the qubit state:

$$\mathcal{N}_{\text{dephase}}(\rho) = (1 - s)\rho + sR_z(\theta)\rho R_z^\dagger(\theta), \quad (12.98)$$

where s is the probability of the noise, and θ is a random variable in $[0, \pi]$.

A more abstract definition that does not specify the random process (or equivalently, averaging over all θ values) leading to the dephasing but focuses on the overall effect is:

$$\mathcal{N}_{\text{dephase}}(\rho) = (1 - s)\rho + s \text{diag}(\rho), \quad (12.99)$$

where $\text{diag}(\rho)$ represents zeroing out off-diagonal elements of ρ and retaining only the diagonal elements.

The complete dephasing channel in Example 12.7 can be considered as the special case of $s = 1$.

5 Amplitude Damping Error

Amplitude damping error models the irreversible decay of the qubit state $|1\rangle$ to the state $|0\rangle$. For example, in a quantum computer, the state $|1\rangle$ typically has a higher energy than the state $|0\rangle$, prompting the system to relax from $|1\rangle$ to $|0\rangle$. The Kraus representation is as follows:

$$\mathcal{N}_{\text{amp-damping}}(\rho) = K_1 \rho K_1^\dagger + K_2 \rho K_2^\dagger, \quad (12.100)$$

where $K_1 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-s} \end{bmatrix}$, $K_2 = \begin{bmatrix} 0 & \sqrt{s} \\ 0 & 0 \end{bmatrix}$, and s is the probability of decay.

Note that amplitude damping error is not a simple random occurrence of two unitary transformations; instead, the probabilities are embedded within the Kraus operators.

6 Pauli Channel

This is a general single-qubit noise channel. Pauli X , Y , and Z matrices serve as Kraus operators to model this channel:

$$\mathcal{N}_{\text{Pauli}}(\rho) = (1 - s_x - s_y - s_z)\rho + s_x X \rho X + s_y Y \rho Y + s_z Z \rho Z, \quad (12.101)$$

where s_x , s_y , and s_z are the probabilities of each Pauli error. Restrictions on these probabilities are not imposed unless specified.

7 Two-qubit Bit-Flip Error

This channel describes the bit-flip process in a two-qubit quantum state. It is an extension of the single-qubit bit-flip error. The Kraus representation of this error channel is given by:

$$\mathcal{N}_{\text{bit-flip-2}}(\rho) = (1 - s)\rho + \frac{s}{3}(IX\rho IX + XI\rho XI + XX\rho XX), \quad (12.102)$$

where s is the strength of the noise. The last three terms represent the flip of the first qubit, the second qubit, and both, respectively. The factor of $\frac{1}{3}$ in the Kraus representation ensures that each of the bit-flip error terms contributes equally to the overall noise process.

8 Two-qubit Dephasing Error

This channel describes the dephasing process in a two-qubit quantum state. It is an extension of the single-qubit phase-flip error. The Kraus representation is:

$$\mathcal{N}_{\text{dephase-2}}(\rho) = (1 - s)\rho + \frac{s}{3}(IZ\rho IZ + ZI\rho ZI + ZZ\rho ZZ), \quad (12.103)$$

where s is the strength of the noise. The factor of $1/3$ in the Kraus representation is conventional and allows for equal contributions from each term.

9 Two-qubit Pauli Channel

This is a general two-qubit noise channel that can model loss of quantum phase, superposition between basis states, and entanglement. The Kraus representation is:

$$\mathcal{N}_{\text{Pauli-2}}(\rho) = \sum_i s_i K_i \rho K_i^\dagger, \quad (12.104)$$

where K_i is one of the sixteen two-qubit Pauli matrices, formed by taking tensor products of single-qubit Pauli matrices, and s_i is the probability of the corresponding K_i .

12.3.3 *Other Types of Errors

Apart from coherent and incoherent errors, there are specific errors that occur during the initialization and readout stages of quantum computation. State preparation and measurement (SPAM) errors are often critical as they can set the stage for or amplify subsequent errors during the computation process.

1 State Preparation Errors

Occurring during the initialization phase, state preparation errors arise when an intended initial state like $|0\rangle$ is realized as a different state such as $|1\rangle$. These errors can be either coherent or incoherent.

State preparation errors can be modeled using a density matrix representation. If the intended initial state is $|\psi\rangle$, the actual state can be described as:

$$\rho_{\text{actual}} = (1 - \varepsilon) |\psi\rangle\langle\psi| + \varepsilon \rho_{\text{error}}, \quad (12.105)$$

where ε is the error rate and ρ_{error} is the density matrix representing the error state.

2 Readout Errors (Measurement Errors)

Readout errors, also known as measurement errors, manifest during the readout process. A state such as $|0\rangle$ might be mis-measured as $|1\rangle$. Similar to state preparation errors, readout errors can also be either coherent, arising from systematic biases, or incoherent, occurring randomly.

Mathematically, readout errors can be modeled using a confusion matrix C , where the element C_{ij} represents the probability of measuring state $|j\rangle$ when the actual state is $|i\rangle$. For a two-level system, the confusion matrix can be expressed as:

$$C = \begin{bmatrix} 1 - \varepsilon_0 & \varepsilon_1 \\ \varepsilon_0 & 1 - \varepsilon_1 \end{bmatrix}, \quad (12.106)$$

where ε_0 and ε_1 are the probabilities of incorrectly reading the states $|0\rangle$ and $|1\rangle$, respectively.

12.3.4 Performance Merit Parameters

Merit parameters in quantum computing provide a quantitative framework to assess the capabilities and limitations of quantum systems. In contrast to classical computers, which are relatively homogenous in technology, quantum computers are developed across various and rapidly advancing hardware platforms. Standardized performance metrics are crucial for comprehensively understanding, comparing, and validating the operations of these diverse quantum technologies.

While the number of qubits is frequently cited as a primary comparison metric, implying that more qubits equate to a more potent system able to manage larger or more complex algorithms, this perspective oversimplifies the intricacies of quantum computation. Given the susceptibility of current quantum computing hardware to noise and consequential error rates that can inhibit performance, a combined assessment of the quantity and the quality of qubits yields a more nuanced evaluation of different quantum systems.

These merit parameters encompass temporal, spatial, and computational dimensions from a physical standpoint. Operationally, they are examined at various levels: qubit, circuit, Quantum Processing Unit (QPU), and the overall system. Furthermore, metrics concerning fault-tolerance and scalability are critical. They offer insights into the long-term practicality and effectiveness of quantum computing architectures, particularly those that aim for fault-tolerant operations and scalable designs.

1 Temporal Metrics

Relaxation Time: T_1

The relaxation time T_1 measures the rate at which an excited qubit returns to its ground state due to energy relaxation, also known as amplitude damping. This parameter determines the effective lifetime of the excited state.

An exponential decay model for the relaxation of the excited state can be represented as:

$$\rho(t) = e^{-t/T_1} \rho(0) + (1 - e^{-t/T_1}) |0\rangle\langle 0|, \quad (12.107)$$

where $\rho(t)$ denotes the system's state at time t . As $t \rightarrow \infty$, the qubit eventually stabilizes in the state $|0\rangle$.

In quantum error correction, the effects of energy relaxation are typically abstracted and modeled as bit-flip errors (§ 12.3.2.1 and § 12.4.1), capturing the transition from excited to ground states. The exponential decay of the excited state corresponds to a constant rate of occurrence for bit-flip errors over time, which can be modeled by a Poisson distribution.

Coherence Time: T_2

The coherence time T_2 quantifies the duration over which a qubit maintains a coherent superposition before dephasing errors transform it into a mixed state. This factor is critical because quantum computation relies on maintaining the phase coherence of qubits for quantum interference and entanglement.

For a generic pure state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the density matrix is:

$$\rho = \begin{bmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{bmatrix}. \quad (12.108)$$

Dephasing can be modeled by an exponential decay of the off-diagonal elements:

$$\rho(t) = \begin{bmatrix} |\alpha|^2 & \alpha\beta^*e^{-t/T_2} \\ \beta\alpha^*e^{-t/T_2} & |\beta|^2 \end{bmatrix}. \quad (12.109)$$

As $t \rightarrow \infty$, the qubit evolves into a mixed state $\rho = |\alpha|^2|0\rangle\langle 0| + |\beta|^2|1\rangle\langle 1|$.

For the purposes of quantum error correction, dephasing effects are commonly represented through phase-flip error channels (§ 12.3.2.2 and § 12.4.2), which model the loss of phase coherence between superposition states.

Qubit Lifetime: T_q

The qubit lifetime T_q is the aggregate period during which a qubit remains in a computationally useful state. Simplistically, T_q can be approximated by $\min(T_1, T_2)$. Typically, T_2 is shorter than T_1 , as phase coherence is susceptible to any environmental fluctuations that can perturb the qubit states, even in the absence of energy exchange.

This metric is affected by various elements, such as the quality of qubit fabrication, the efficiency of error correction strategies, and the specific quantum computing architecture employed. Different qubit technologies exhibit a wide range of gate lifetimes and other key temporal parameters. For instance, superconducting qubits often display qubit lifetimes ranging from microseconds to milliseconds. In contrast, systems based on optical qubits, like trapped ions, can sustain coherence for significantly longer durations, spanning from seconds to minutes in some cases.

Gate Time: T_{gate}

The operational speed of a QPU is intrinsically related to the duration of its gate operations. The gate time T_{gate} is the time required to execute a single quantum gate operation.

Coherence Time to Gate Time Ratio

The ratio of coherence time T_q to gate time T_{gate} holds practical significance. If T_{gate} is considerably shorter than T_q , a larger number of successive gate operations can be executed before coherence is lost. Consequently, the achievable depth of quantum circuits on a QPU is generally limited by this ratio.

For instance, while optical systems may exhibit greater coherence times, their control pulse durations are also longer. This means that the ratio of qubit coherence to gate operation times may not necessarily be more advantageous than that of superconducting or semiconducting systems, which operate with pulses on the nanosecond (ns) timescale.

2 Spatial Metrics

Connectivity or Topology

Connectivity, or topology, in quantum computing describes the manner in which qubits within a system are interconnected, facilitating quantum operations between them. This aspect is important in determining how easily multi-qubit gates can be executed.

High connectivity eases the implementation of quantum algorithms by reducing the need for additional SWAP gates, which are otherwise necessary to bring non-adjacent qubits into interaction range. However, increased connectivity can also complicate the physical design of the system and introduce additional error sources, such as cross-talk.

Connectivity types range from nearest-neighbor (where qubits interact only with adjacent ones) to all-to-all (permitting direct interaction between any pair of qubits). The connectivity pattern is closely tied to the quantum computer's architecture and

shows significant differences across technologies. For example, solid-state devices like superconducting circuits and semiconducting qubits are typically limited in connectivity due to the constraints of physical wiring or resonators. These systems often have just two or three directly connected neighbors. Entangling distant qubits in such architectures requires additional operations, like SWAP gates or teleportation, to transfer quantum information.

In contrast, trapped ion and optical systems are not constrained by a two-dimensional circuit layout, thereby offering greater flexibility in qubit addressing and entanglement. These technologies can often achieve an all-to-all connectivity, facilitating more direct and versatile qubit interactions for algorithm execution and logic processing. However, it is important to note that the physical movement of ions necessary for these interactions in trapped ion systems can be time-consuming, potentially impacting the overall operation speed of the quantum computer.

Cross-Talk

Cross-talk in quantum computing refers to the inadvertent interactions between qubits that are not intended to be coupled at a particular moment. Such interference typically stems from electromagnetic disturbances, stray couplings, or leakage in control signals.

This phenomenon can inadvertently alter the states of qubits not actively participating in a specific operation, potentially leading to errors and reduced fidelity in quantum computations. To combat cross-talk, strategies such as meticulous design of the physical layout, shielding, implementation of decoupling sequences, and precise calibration routines are utilized. Advanced control methods are also employed for dynamic compensation of cross-talk effects.

As quantum systems scale up, understanding and mitigating cross-talk becomes increasingly vital for developing reliable and fault-tolerant quantum circuits. The challenge lies in achieving high connectivity while simultaneously maintaining qubit coherence and minimizing cross-talk and other error sources in quantum computer design.

3 Computational Metrics

Fidelity

Fidelity quantifies the precision with which a quantum gate or circuit performs in comparison to its ideal theoretical model. A fidelity value of 1 signifies flawless execution, whereas values less than 1 reflect discrepancies from the ideal operation.

The expected output of an ideal quantum gate or circuit is typically a pure state, denoted as $|\psi_{\text{ideal}}\rangle$. However, due to noise and other factors, the actual output might be a mixed state, represented by ρ_{actual} . The fidelity in this scenario is defined as:

$$F = \langle \psi_{\text{ideal}} | \rho_{\text{actual}} | \psi_{\text{ideal}} \rangle. \quad (12.110)$$

This formulation encompasses both the desired unitary operations and the non-unitary effects of noise. For cases where the actual output is also a pure state $|\psi_{\text{actual}}\rangle$, fidelity simplifies to $F = |\langle \psi_{\text{ideal}} | \psi_{\text{actual}} \rangle|^2$.

Fidelity can be assessed at both the gate and circuit levels. Generally, circuit fidelity tends to decrease with increasing circuit depth due to error accumulation, as expressed by:

$$F \approx F_G^D, \quad (12.111)$$

where F is the overall circuit fidelity, F_G represents the average gate fidelity, and D is the circuit depth.

Notably, the fidelity of two-qubit gates, such as CNOT or ZZ, frequently determines the upper limit of the system's overall fidelity. This is because two-qubit gates generally present more challenges in achieving high precision.

Error Rate

Error rate quantifies the probability of an error occurring during a quantum operation, which could be a single gate or a sequence of gates. Techniques such as randomized benchmarking are often used to estimate this rate.

Fidelity (F) and error rate (ε) are complementary metrics. They are related by:

$$\varepsilon = 1 - F. \quad (12.112)$$

Circuit Depth

Circuit depth refers to the maximum number of sequential gate operations that can be performed before the system accrues an unacceptable level of errors. As a rough approximation, circuit depth (D) and the gate-level error rate (ε_G) are inversely related:

$$D \approx \frac{1}{\varepsilon_G}, \quad (12.113)$$

indicating that higher error rates limit the feasible depth of quantum circuits.

4 System-Level Composite Metrics

In the rapidly evolving field of quantum computing, single-number composite metrics offer a holistic approach to evaluate the performance of quantum computing systems. These metrics aim to distill the complex, multi-faceted characteristics of quantum systems into a single, comprehensive figure, facilitating a more straightforward comparison and understanding of quantum processors from an application-level perspective. While beneficial for providing an overarching view of a quantum computer's capabilities, these metrics are continually evolving, reflecting the dynamic nature of quantum technology and its diverse applications. Below are a few representative examples.

Quantum Volume

Quantum Volume (QV), introduced by IBM, serves as a composite metric designed to provide a unified framework for evaluating and comparing quantum processors. It accounts for factors such as the number of qubits, coherence, gate fidelity, and connectivity.

QV is determined by the size of the largest square circuit that can be successfully run on a QPU. The term "square" refers to a circuit where the number of qubits (width) is equal to the number of layers of two-qubit gates (depth). The protocol for determining QV involves running a series of circuits that increase in complexity, both in terms of width and depth, and identifying the largest circuit that the quantum computer can reliably execute. These circuits consist of layers of qubit permutation

(SWAP) and random single and double qubit gates. If the number of qubits in the largest successfully executed square circuit is n , then QV is given by 2^n .

QV, being dependent on both circuit depth and width, inherently relates QPU performance to its topology, error rates, and overall size. As such, QV is increasingly used as a gauge of QPU performance, offering a holistic and comprehensive view through a single value.

CLOPS

While Quantum Volume (QV) provides insight into the potential capabilities of a quantum computer, it does not fully capture the operational speed at a system level. This is particularly relevant for quantum-classical interfacing in full applications, where algorithms require multiple calls to a QPU. The efficiency of the runtime system, which facilitates quantum-classical communication, is crucial for high performance.

CLOPS, or Circuit Layer Operations Per Second, measures the speed and capability of a quantum computer to solve problems of a certain complexity. It defines how many QV circuits a quantum processor can process within a given timeframe, thus providing an operational speed measure in practical scenarios. CLOPS emphasizes the practical throughput of executing quantum algorithms, taking into account the efficiency of quantum-classical communication and the runtime system.

A pertinent metric in this context is “teraquops”, the quantum counterpart of “teraops” in classical computation. It signifies the number of quantum operations (quops) a quantum system is capable of performing each second, scaled to a tera-order of magnitude (trillions of operations per second). This metric gauges the raw computational power and speed of quantum systems, complementing the throughput-centric perspective provided by CLOPS.

Algorithmic Qubits

Quantum Volume (QV) is a widely recognized metric, yet there is a growing consensus that it may not fully encompass the nuances of quantum computational power, especially for specific algorithms or practical use cases. Concerns have been raised that QV might not accurately reflect the performance of quantum computers in scenarios that do not require deep circuits or extensive qubit connectivity. Furthermore, QV may not represent the capabilities of various quantum computing platforms, such as quantum annealers or trapped ion systems, as effectively as it does for superconducting qubit systems.

In response, the concept of Algorithmic Qubits (AQ) has been proposed as an alternative metric. Unlike QV, which assesses a QPU’s ability to execute increasingly complex quantum circuits, AQ focuses on the “algorithmic strength” of a quantum computer by considering error rates, qubit coherence, gate fidelity, and the overall system architecture in relation to specific quantum algorithms.

AQ provides valuable insights into the practical usability of a quantum processor for executing complex algorithms, offering an application-oriented perspective. It indicates the number of qubits that can be effectively used for algorithmic tasks, factoring in error correction and mitigation techniques necessary for meaningful computational outcomes. AQ thus complements QV, aiding in the selection of the most appropriate quantum processor for specific algorithmic requirements.

5 Fault-Tolerance and Scalability Metrics

Fault-tolerance and scalability metrics are instrumental in providing insights into the viability and efficiency of fault-tolerant quantum computing architectures. By analyzing these metrics, researchers and engineers are equipped to identify optimal architectures and error correction strategies, thereby facilitating the practical realization of robust and high-performance quantum computers.

These metrics primarily encompass an analysis of noise structure, the relationship between physical noise levels and fault-tolerant thresholds (quantified by the Λ factor), the scaling properties of quantum error-correcting codes (QECC), the complexity of QECC decoding, and the computational resource requirements essential for achieving fault tolerance. The specific metrics of interest include:

The Λ Factor

The Λ factor is an essential metric, defined as the ratio of the threshold error rate for a given QECC, $\varepsilon_{\text{threshold}}$, to the actual physical error rate of quantum operations, $\varepsilon_{\text{physical}}$:

$$\Lambda = \frac{\varepsilon_{\text{threshold}}}{\varepsilon_{\text{physical}}}.$$
 (12.114)

A higher Λ factor is indicative of physical error rates being substantially below the threshold, thereby facilitating effective error correction and extending coherence times within the system.

Application Footprint

The application footprint of a specific quantum computing architecture is a measure that quantifies the resource requirements for a quantum computing task. It is gauged in terms of the number of physical qubits necessary to realize a certain number of logical qubits with a target error rate (e.g., 10^{-12}) or to address a particular computational problem (e.g., factoring a 2048-bit integer).

This metric elucidates the scalability of a quantum computing architecture, underscoring the trade-offs between the overhead of error correction and computational capability. It is invaluable for evaluating the feasibility of employing large-scale quantum algorithms for real-world applications.

Subthreshold Scaling

Subthreshold scaling is a metric that examines the performance dynamics of a QECC as the physical error rate falls below the threshold error rate. This analysis yields insights into the efficiency and dependability of the error correction code, particularly as the system's size and complexity expand. This metric is crucial for assessing the long-term sustainability and performance viability of QECCs.

QECC Decoding Complexity

The complexity of decoding QECCs varies significantly among different types. For instance, surface codes demand minimal additional processing for error correction. In contrast, Low-Density Parity-Check (LDPC) codes necessitate extensive computational resources for real-time decoding, which presents challenges for scalability. This measure of decoding complexity serves as an indicator of the resources required—specifically in terms of time, memory, and classical computational effort—to effectively identify and correct errors in a quantum computing system.

Capability to Suppress a Broad Spectrum of Errors

Some errors, particularly coherent errors such as those encountered in state preparation, gate calibration, and measurement, cannot be effectively corrected by current QECCs. Nevertheless, for a quantum computing system to scale effectively, it is crucial to develop strategies capable of suppressing a comprehensive range of errors that undermine computational accuracy.

12.4 Introduction to Error Correction Codes

Achieving fault tolerance is an essential yet elusive goal for the advancement of quantum computing. Theoretically, full fault tolerance—often considered the Holy Grail in the field—becomes attainable once the actual error rate falls below a certain threshold. However, this comes at the expense of requiring a significant qubit overhead, making effective error correction a primary research focus.

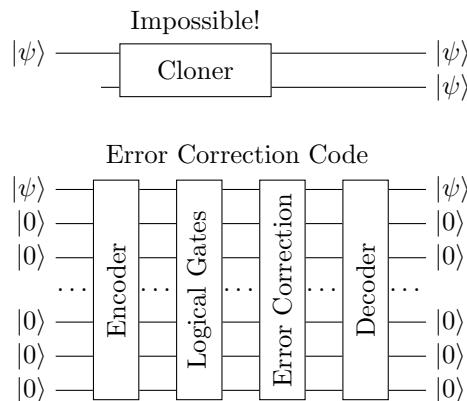


Figure 12.11: Structure of Practical Error Correction Code

Quantum error correction follows a systematic procedure known as an error correction code. As illustrated in Fig. 12.11, this procedure encompasses the encoding of logical qubits into a larger set of physical qubits, the detection and identification of errors, and the application of corrective operations. Subsequently, decoding is performed to retrieve the corrected logical information. The ultimate objective is to approximate a nearly perfect logical qubit by employing a sufficiently large array of physical qubits.

In this section, we introduce the fundamentals of quantum error correction codes through the study of Shor codes, an archetypal example in the field. Shor codes use a blend of quantum entanglement, local operations, and principles from classical error correction to bolster the fidelity of quantum information processing. Beyond its utility in fault-tolerant quantum computing, Shor codes also find applications in secure quantum communications.

We will focus on the comprehensive nine-qubit Shor code. To fully grasp the sophistication behind Shor's code, we start by examining two simpler but foundational



Peter Shor

Source: International Centre
for Theoretical Physics

codes: the three-qubit bit-flip code and the three-qubit phase-flip code. Shor's code is a masterful combination of these elementary codes, enabling it to correct both bit-flip and phase-flip errors simultaneously, and thus effectively addressing any one-qubit error.

12.4.1 Three-Qubit Bit-Flip Code

The three-qubit bit-flip code is a quantum error correction code designed to guard against single bit-flip errors. The quantum circuit diagram for this code is depicted in Fig. 12.12.

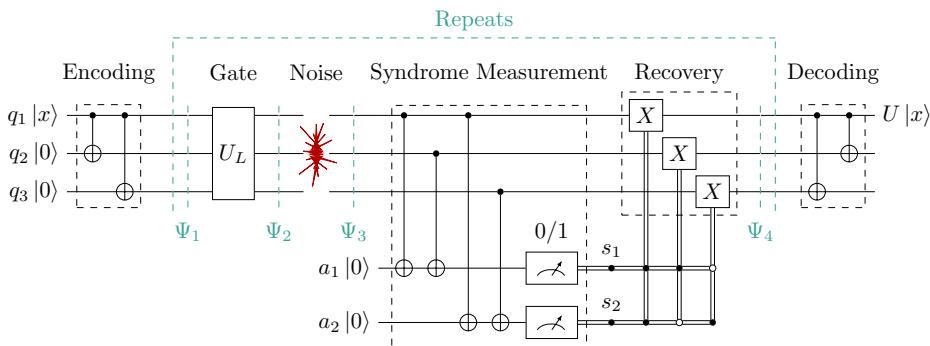


Figure 12.12: Quantum Circuit for the Three-Qubit Bit-Flip Code

Computation in Place

In quantum circuits, gates do not exist simultaneously as they might on a classical printed circuit board or integrated circuit. Contrarily, in the quantum realm, neither gates nor wires manifest as palpable physical entities. What's physically implemented are only the qubits and their corresponding external controls. In this context, the wires represent the temporal continuity of the qubit states, and the quantum gates denote specific transformations of these qubit states, directed by the external controls.

For a more information, refer to § 5.3.2.

1 Encoding

The encoding process shown in Fig. 12.12 utilizes three data qubits q_1 , q_2 , and q_3 and two CNOT gates. In this context, $|x\rangle$ represents the basis states $|0\rangle$ and $|1\rangle$. After encoding, it becomes either $|0_L\rangle \equiv |000\rangle$ or $|1_L\rangle \equiv |111\rangle$. A generic single qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is mapped to a logical qubit state as:

$$|\Psi_1\rangle = |\psi_L\rangle \equiv \alpha|0_L\rangle + \beta|1_L\rangle \equiv \alpha|000\rangle + \beta|111\rangle. \quad (12.115)$$



It's essential to note that $|\psi_L\rangle \neq |\psi\psi\psi\rangle$. This means that while $|\psi\rangle$ spreads across three qubits via entanglement, it isn't replicated thrice. Making such copies would violate the no-cloning theorem.

2 Logical Gates

When employing an error-correction code, qubit gates, such as the Pauli X and Z , must be translated into logical gates that act on the encoded logical state. These logical gates may affect multiple data qubits, but their operation does not take the system out of the code space, i.e., the space spanned by the logical basis states $\{|000\rangle, |111\rangle\}$. Consider a unitary gate U :

$$U(\alpha|0\rangle + \beta|1\rangle) = \gamma|0\rangle + \delta|1\rangle. \quad (12.116)$$

The logical U gate, given by U_L , transforms the state of the system into:

$$|\Psi_2\rangle = \gamma|000\rangle + \delta|111\rangle. \quad (12.117)$$

Exercise 12.22 Show that $X_L \equiv XXX$ functions as the logical bit-flip operator for the three-qubit bit-flip code.

This task exemplifies the concept of *transversal gates*, which involve applying physical gates individually to each qubit of a quantum error-correcting code word, without necessitating multi-qubit operations that could potentially propagate errors.

3 Noise Channel

The gate operation of U_L can be a long, complex process during which errors might occur. Bit-flip errors can be modeled as random applications of the X gates on data qubits. For instance, applying X to q_2 (or equivalently, IXI to all three qubits) flips the bit in q_2 . This noise alters the state of the data qubits to:

$$|\Psi_3\rangle = \gamma|010\rangle + \delta|101\rangle. \quad (12.118)$$

4 Syndrome Measurement

The syndrome measurement circuit identifies any modifications to the qubits resulting from noise. It employs two ancilla qubits. The first, a_1 , checks the parity of data qubits q_1 and q_2 with two CNOT gates. If q_1 and q_2 share the same state, either $|0\rangle$ or $|1\rangle$, the dual CNOT gates negate each other, and the measurement produces $s_1 = 0$. Thus, in the absence of noise, the state $\gamma|000\rangle + \delta|111\rangle$ invariably results in $s_1 = 0$, regardless of γ and δ values.

If a bit flip in q_2 has occurred, q_1 and q_2 now have contrasting states, and the measurement on a_1 returns $s_1 = 1$, highlighting the core principle of parity measurement which discerns the parity of two data qubits without perturbing their states.

In a similar vein, the second ancilla, a_2 , measures the parity of data qubits q_1 and q_3 , generating a classical bit s_2 . This parity information, represented by the ancilla outputs $\{s_1, s_2\}$, is termed the error syndrome. Table 12.1 summarizes the possible error syndromes, corresponding errors, and their corrective measures.

$s_{1,2}$	Error	Correction Gates
00	No error	None
01	Bit flip on q_3	X on q_3
10	Bit flip on q_2	X on q_2
11	Bit flip on q_1	X on q_1

Table 12.1: Error Syndrome and Correction for the Three-qubit Bit-flip Code

5 Error Correction and State Recovery

Upon detecting the error syndrome, the error can be rectified by applying an X gate to the relevant qubit as specified by Table 12.1. This action flips the aberrant qubit back to its noise-free state. Now the system state is recovered:

$$|\Psi_4\rangle = |\Psi_2\rangle = \gamma|000\rangle + \delta|111\rangle. \quad (12.119)$$

Utilizing the deferred measurement principle, as detailed in § 7.4.5, the classically controlled X gates can be substituted with CCNOT gates.

6 Sequential Gate Operations

A quantum computation task is typically carried out with a sequence of gates. As errors accumulates above a certain threshold, error detection and state recovery are executed. This is represented as repeating the gate-noise-syndrome-recovery block in Fig. 12.12.

7 Decoding

The decoder's role is to transform the logical qubit state $\gamma|000\rangle + \delta|111\rangle$ back to $\gamma|0\rangle + \delta|1\rangle$, which is $U|\psi\rangle$. This is achieved by applying two CNOT gates in the inverse sequence of the encoder.

8 Undetected Errors

It is important to address certain limitations of the Shor three-qubit bit-flip code. Referring to Table 12.1, we have made the assumption that only a single qubit experiences a bit-flip error. However, the syndrome 01 can be a result of bit-flip errors on both q_1 and q_2 , corresponding to the error operator XXI . Similarly, the syndrome 11 can emerge from the error operator IXX , and the syndrome 00 might be due to XXX . Such double or triple bit-flip errors, although theoretically possible, are less frequent. Their probability is of the order $O(\varepsilon^2)$ or lower, making them less likely when compared to single bit-flip errors, especially when the single qubit error rate ε is sufficiently low.

Moreover, the Shor three-qubit bit-flip code is specifically designed to counteract bit-flip errors. Consequently, errors involving phase flips, represented by IIZ , IZI , or ZII , are not detected. Each of these transforms the state $\gamma|000\rangle + \delta|111\rangle$ to $\gamma|000\rangle - \delta|111\rangle$, with the parity unaffected. This limitation underscores the necessity of the phase-flip code.

Exercise 12.23 Reflect on the potential benefits of incorporating a third ancilla to measure the parity of q_2 and q_3 . Could it enhance the capability to detect more types of errors?

Exercise 12.24 Investigate if the Shor three-qubit bit-flip code can detect and correct errors that result from random applications of Y gates. Aim to construct a table akin to Table 12.1 and an error correction code analogous to Fig. 12.12.

12.4.2 Three-Qubit Phase-Flip Code

The Shor phase-flip code is similar to its bit-flip counterpart, as illustrated in Fig. 12.13. The application of a Hadamard rotation transforms bit-flip operations to correct phase flips.

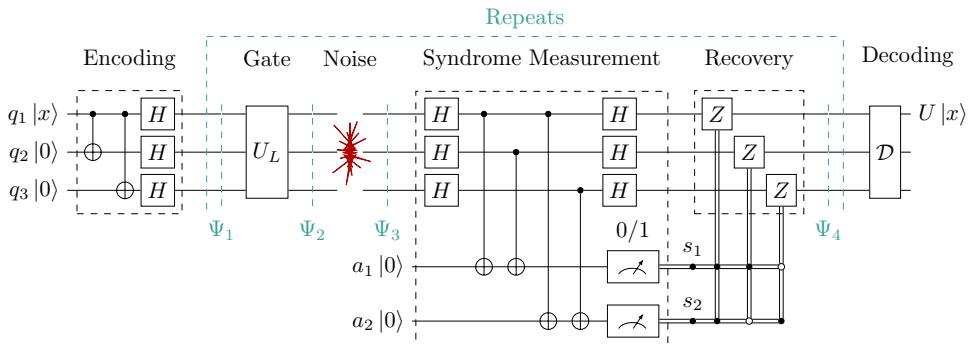


Figure 12.13: Quantum Circuit for the Three-Qubit Phase-Flip Code

The Three-Qubit Phase-Flip Code maps the initial qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ to a three-qubit state as follows:

$$|\Psi_1\rangle = |\psi\rangle_L \equiv \alpha|+_L\rangle + \beta|_-L\rangle \equiv \alpha|+++> + \beta|--->, \quad (12.120)$$

where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.

After the application of the logical gate U_L , the system state becomes:

$$|\Psi_2\rangle = \gamma|+++> + \delta|--->. \quad (12.121)$$

Exercise 12.25 Show that for the three-qubit phase-flip code, ZZZ functions as the logical bit-flip operator X_L .

A phase-flip error can be modeled as random applications of Z gates on data qubits, since $Z|+\rangle = |-\rangle$ and $Z|-\rangle = |+\rangle$. For instance, applying Z to q_1 (equivalent to ZII) flips the phase in q_1 . This noise alters the state of the data qubits to:

$$|\Psi_3\rangle = \gamma|---> + \delta|++->. \quad (12.122)$$

Using the Hadamard gates, we can translate $\gamma|-\!+\!+\rangle + \delta|+\!-\!-\rangle$ into $\gamma|100\rangle + \delta|011\rangle$ and revert it. This process enables the CNOT gates to detect any noise-induced modifications to the qubits. For instance, in our scenario, the error manifests as a syndrome 11.

Table 12.2 summarizes the possible error syndromes, corresponding errors, and their corrective measures.

$s_{1,2}$	Error	Correction Gates
00	No error	None
01	Phase flip on q_3	Z on q_3
10	Phase flip on q_2	Z on q_2
11	Phase flip on q_1	Z on q_1

Table 12.2: Error Syndrome and Correction for the Three-Qubit Phase-Flip Code

The error can be rectified by applying a Z gate to the affected qubit (in our scenario, q_1), restoring the system state to:

$$|\Psi_4\rangle = |\Psi_2\rangle = \gamma|+++\rangle + \delta|---\rangle. \quad (12.123)$$

The decoder circuit, represented by \mathcal{D} in Fig. 12.13, mirrors the encoder circuit. It decodes $|\Psi_4\rangle$ back to $\gamma|0\rangle + \delta|1\rangle$, equivalent to $U|\psi\rangle$.

Similar to the three-qubit bit-flip code, the three-qubit phase-flip Code is not equipped to detect double or triple phase-flip errors, nor any bit-flip errors.

Exercise 12.26 Modify the circuit in Fig. 12.13 for the three-qubit phase-flip code to employ classically controlled X gates instead of classically controlled Z gates for error correction.

12.4.3 * Nine-Qubit Shor Code

The nine-qubit Shor code, introduced by Peter Shor in 1995, represents one of the pioneering quantum error-correcting codes. This code is devised by concatenating the three-qubit bit-flip and three-qubit phase-flip codes, as demonstrated in Fig. 12.14. This construction allows for the correction of any single-qubit errors, which will be elaborated upon subsequently.

1 Encoding

The encoding process integrates both the phase-flip and bit-flip codes, encoding the logical qubits of the phase-flip code using the bit-flip code.

The initial CNOT and Hadamard gates perform phase-flip code encoding on the qubits q_1 , q_4 , and q_7 . Following this first-level encoding, the state of the entire nine-qubit system evolves as:

$$|0\rangle \mapsto \frac{1}{\sqrt{8}}(|0\rangle + |1\rangle)|00\rangle \quad (|0\rangle + |1\rangle)|00\rangle \quad (|0\rangle + |1\rangle)|00\rangle, \quad (12.124a)$$

$$|1\rangle \mapsto \frac{1}{\sqrt{8}}(|0\rangle - |1\rangle)|00\rangle \quad (|0\rangle - |1\rangle)|00\rangle \quad (|0\rangle - |1\rangle)|00\rangle. \quad (12.124b)$$

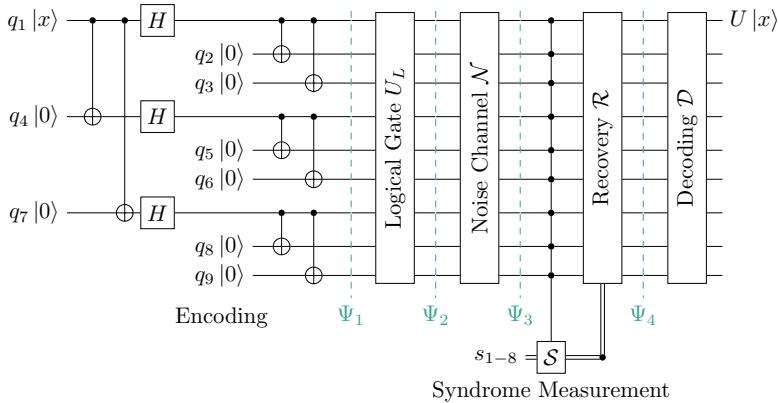


Figure 12.14: Quantum Circuit for the Nine-Qubit Shor Code

The CNOT gates that follow execute bit-flip code encoding on qubits q_1 through q_9 . This operation transitions the basis states to their corresponding logical states:

$$|0\rangle \mapsto |0_L\rangle = \frac{1}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle), \quad (12.125a)$$

$$|1\rangle \mapsto |1_L\rangle = \frac{1}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle). \quad (12.125b)$$

A quantum state generally represented as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is encoded into:

$$|\Psi_1\rangle = |\psi_L\rangle = \alpha|0_L\rangle + \beta|1_L\rangle. \quad (12.126)$$

Exercise 12.27 Elucidate the reason that the encoding procedure, as presented in Fig. 12.14, does not directly result in $|0\rangle \mapsto |+++++++\rangle$ and $|1\rangle \mapsto |-----\rangle$, but rather adheres to the transformation depicted in Eq. 12.125. Reference: § 7.3.3.

2 Logical Gates

When employing an error-correction code, qubit gates must be translated into the corresponding logical gates (i.e., logical operations) that act on the encoded logical states. Consider a unitary gate U :

$$U(\alpha|0\rangle + \beta|1\rangle) = \gamma|0\rangle + \delta|1\rangle. \quad (12.127)$$

The logical U gate, denoted by U_L , is to be compatible with the error correction code, meaning it transforms a logical basis state to another logical basis state or a superposition of such states. After applying U_L , the system state is given by:

$$|\Psi_2\rangle = \gamma|0_L\rangle + \delta|1_L\rangle \quad (12.128a)$$

$$\begin{aligned} &= \frac{\gamma}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle) \\ &\quad + \frac{\delta}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle). \end{aligned} \quad (12.128b)$$

Designing logical gates poses significant challenges. The following exercises offer an introductory exploration. The Clifford group, which includes Pauli gates (X , Y , Z), the Hadamard gate (H), the phase gate (S), and the CNOT gate, can be efficiently implemented with fault tolerance in various quantum error-correcting codes. For more discussion, refer to § 7.2.2.

Exercise 12.28 The operator $Z_1Z_4Z_7$ interchanges the amplitudes of the logical qubit states, effectively acting as a logical X gate (X_L). Demonstrate how applying $Z_1Z_4Z_7$ to the state $\alpha|0_L\rangle + \beta|1_L\rangle$ results in $\alpha|1_L\rangle + \beta|0_L\rangle$.

Similarly, a logical Z gate (Z_L) modifies the phase between the logical states. Show that applying $Z_L = X_1X_2X_3$ to $\alpha|0_L\rangle + \beta|1_L\rangle$ transforms it into $\alpha|0_L\rangle - \beta|1_L\rangle$, thus serving as a logical Z gate.

The nine-qubit Shor code supports multiple implementations of both X_L and Z_L . Identify two additional operators for each that fulfill the logical X and Z functions, respectively.

3 Error Syndrome Measurement and Correction

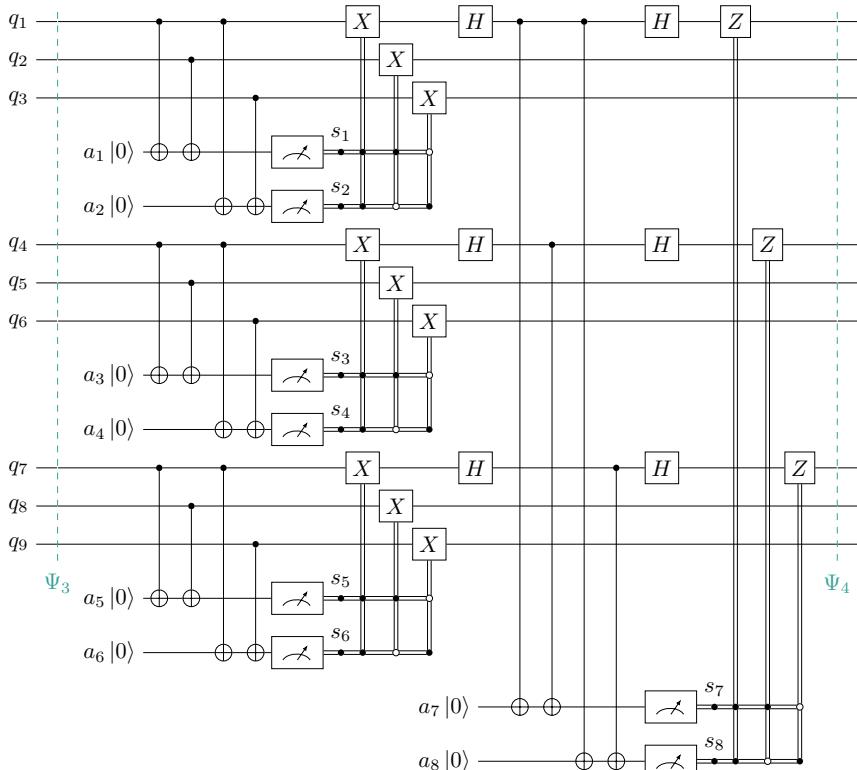


Figure 12.15: Error Syndrome Measurement and Correction for the Shor Code

As shown in Fig. 12.15, the nine-qubit Shor code utilizes three blocks of the three-qubit bit-flip code. Each block can detect and correct a bit-flip error within its group of three qubits. Similarly, for phase errors, the Shor code employs the three-qubit phase-flip code on three encoded qubits, one from each of the three

blocks. This structure, combined with the associated Hadamard gates, can detect which block (if any) had a phase-flip error and correct it.

We will examine a few typical scenarios with examples next.

4 Bit-Flip Error

When a bit-flip error occurs, for instance, on q_5 , the system state becomes:

$$|\Psi_3\rangle = \frac{\gamma}{\sqrt{8}}(\dots)(|010\rangle + |101\rangle)(\dots) + \frac{\delta}{\sqrt{8}}(\dots)(|010\rangle - |101\rangle)(\dots). \quad (12.129)$$



For clarity, in the preceding equation and in subsequent discussions, the changed bit is highlighted, and the unchanged blocks are represented using (\dots) .

This bit-flip error is detected by the second block of the bit-flip code, yielding the error syndrome $s_3s_4 = 10$. Consequently, the corrective X gate is applied to q_5 , returning the system state to $|\Psi_4\rangle = |\Psi_2\rangle$.

5 Phase-Flip Error

When a phase-flip error takes place, for instance, on q_5 , the system state evolves to:

$$|\Psi_3\rangle = \frac{\gamma}{\sqrt{8}}(\dots)(|000\rangle - |111\rangle)(\dots) + \frac{\delta}{\sqrt{8}}(\dots)(|000\rangle + |111\rangle)(\dots). \quad (12.130)$$

This phase-flip error is identified by the phase-flip code, generating the error syndrome $s_7s_8 = 10$. As a result, the corrective Z gate is applied to q_4 , bringing the system state back to $|\Psi_4\rangle = |\Psi_2\rangle$.

Note that the Shor code cannot distinguish which of the qubits q_4 , q_5 , or q_6 has experienced a phase-flip. This inability to differentiate is referred to as error degeneracy, a phenomenon where different errors yield the same syndrome. However, this lack of specificity in pinpointing the exact qubit does not hinder the ability to correct the error. The Shor code can effectively correct the phase flip without needing to know which of the qubits was affected.

6 General Single-Qubit Error

Consider a single-qubit error acting on q_5 such that:

$$|0\rangle \rightarrow a|0\rangle + b|1\rangle, \quad (12.131a)$$

$$|1\rangle \rightarrow c|0\rangle + d|1\rangle. \quad (12.131b)$$

The resulting system state can be represented as:

$$\begin{aligned} |\Psi_3\rangle &= \frac{\delta}{\sqrt{8}}(\dots)(a|000\rangle + b|010\rangle + c|101\rangle + d|111\rangle)(\dots) \\ &\quad + \frac{\gamma}{\sqrt{8}}(\dots)(a|000\rangle + b|010\rangle - c|101\rangle - d|111\rangle)(\dots). \end{aligned} \quad (12.132)$$

To categorize bit-flip and phase-flip errors, we introduce the following relationships: $p+q = a$, $p-q = d$, $u+v = b$, and $u-v = c$. The terms are then rearranged as outlined in Table 12.3.

Term in $ \Psi_3\rangle$	$s_{3,4,7,8}$	Error
$\frac{p}{\sqrt{8}}(\dots)(\delta(000\rangle + 111\rangle) + \gamma(000\rangle - 111\rangle))(\dots)$	0000	No error
$\frac{q}{\sqrt{8}}(\dots)(\delta(000\rangle - 111\rangle) + \gamma(000\rangle + 111\rangle))(\dots)$	0010	Phase flip
$\frac{u}{\sqrt{8}}(\dots)(\delta(010\rangle + 101\rangle) + \gamma(010\rangle - 101\rangle))(\dots)$	1000	Bit flip
$\frac{v}{\sqrt{8}}(\dots)(\delta(010\rangle - 101\rangle) + \gamma(010\rangle + 101\rangle))(\dots)$	1010	Bit & phase flip

Table 12.3: Error Syndrome for the Nine-Qubit Shor Code

As delineated in Table 12.3, a general single-qubit error can manifest in one of four scenarios:

1. No error.
2. Bit flip, corrected using an X gate.
3. Phase flip, corrected using a Z gate.
4. Both bit and phase flip, corrected using both X and Z gates.

Exercise 12.29 Construct an error syndrome and correction table similar to Table 12.3, assuming a general single-qubit error has occurred in qubit q_7 , instead of q_5 . For each scenario, identify the appropriate correction gate(s) in Fig. 12.15.

A pertinent question arises: how is it possible for a general error characterized by continuous coefficients a, b, c , and d to be effectively corrected using only four discrete scenarios? The answer is rooted in the fundamental properties of quantum measurements. When one conducts bit- and phase-flip parity-check measurements on the system, it collapses the system's state to one wherein either a bit or phase flip has transpired, or it hasn't, contingent on the outcome of the measurement. This inherent quantum behavior equips us with the capability to address a spectrum of errors merely by executing bit- and phase-flip checks.

Alternatively, we can analyze nine-qubit Shor code using a Pauli noise channel (see § 12.3.2), which is expressed as:

$$\mathcal{N}_{\text{Pauli}}(\rho) = (1 - s_x - s_y - s_z)I\rho I + s_x X\rho X + s_y Y\rho Y + s_z Z\rho Z, \quad (12.133)$$

where s_x , s_y , and s_z are the probabilities of each Pauli error. We find that scenario (1) above corresponds to I in the Pauli channel, scenario (2) to X , (3) to Z , and (4) to Y .

7 Decoding

The decoder's role is to transform the logical qubit state $\gamma|0_L\rangle + \delta|1_L\rangle$ back to $\gamma|0\rangle + \delta|1\rangle$, which is $U|\psi\rangle$. This is achieved by applying the CNOT gates and H gates in the reverse order of the encoder.

8 Sequential vs. Oneshot Operation

A quantum computation task is typically carried out with a sequence of gates. As errors accumulate above a certain threshold, error detection and state recovery are

executed. The version of the Shor code illustrated in Fig. 12.14 is capable of this purpose by repeating the blocks between $|\Psi_2\rangle$ and $|\Psi_4\rangle$.

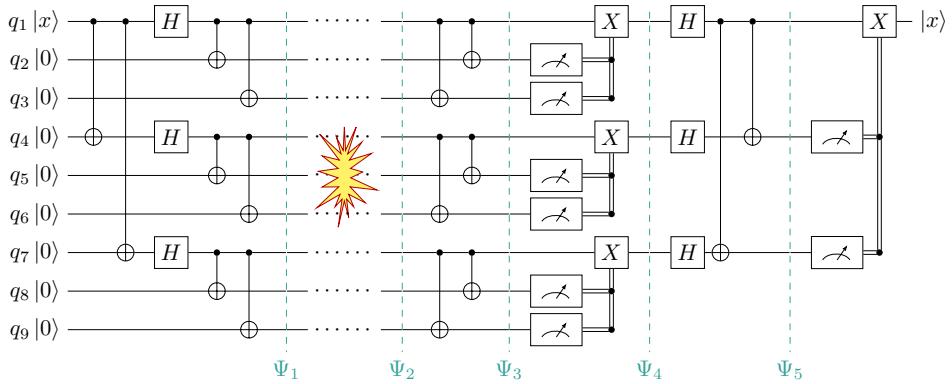


Figure 12.16: Nine-Qubit Shor Code Simplified for Oneshot Operation

The Shor code can also be used for quantum communication in which quantum states are transmitted subject to noise. In this case, the repetition of the middle blocks is often unnecessary, and the circuit can be simplified, as illustrated in Fig. 12.16. In particular, the error correction process can be postponed until after the decoding, which eliminates the need for ancillary qubits.

9 Limitations

We have seen that the Shor code can correct an arbitrary error on a single qubit. When considering simultaneous multi-qubit errors, the situation becomes more complex. Nevertheless, these scenarios are less frequent. Their probability is on the order of $O(\varepsilon^2)$, making them less probable compared to single bit-flip errors, especially when the single qubit error rate ε is sufficiently small.

While repetition codes, such as the Shor code, elucidate the principle of error correction, they are often too inefficient for practical application. This has motivated the search for more efficient codes. A notable example is the seven-qubit Steane code. To delve deeper into such error correction codes, additional mathematical tools, like the stabilizer formalism, are required. These will be introduced in the subsequent subsection.

Exercise 12.30 For the simplified version of the nine-qubit Shor code depicted in Fig. 12.14, derive the system state at different stages ($|\Psi_1\rangle$ through $|\Psi_5\rangle$) for each type of single-qubit error: bit-flip, phase-flip, and general errors.

12.4.4 * Parity Measurements, Stabilizers, and Code Distance

Fundamental to the study of quantum error-correction codes are the inter-related concepts of parity measurements, stabilizer codes, and code distance.

1 Parity Measurements

Parity measurements constitute a type of joint measurement on multiple qubits, as discussed in § 6.2.2. For a two-qubit system, $P_Z \equiv Z \otimes Z$ can distinguish

between the even parity states $|\psi_{\text{even}}\rangle \equiv \alpha|00\rangle + \beta|11\rangle$ and odd parity states $|\psi_{\text{odd}}\rangle \equiv \gamma|01\rangle + \delta|10\rangle$. Similarly, $P_X \equiv X \otimes X$ can be used to measure phase parity.

In the discussion of the Shor code in preceding sections, we have employed parity measurements. Specifically, P_Z is used for the error syndrome measurement associated with bit flips, while P_X facilitates the error syndrome measurement related to phase flips.

The stabilizer formalism discussed next can be viewed as a generalization of the concept of parity measurements.

2 The Stabilizer Formalism

The stabilizer formalism is a framework that enables the efficient description and manipulation of quantum states and operations pertinent to quantum error correction. Rather than representing these entities in Hilbert space, the formalism treats them in the operator picture as elements of a stabilizer group. This framework proves essential in analyzing advanced error correction codes.

Stabilizers are a set of operators that commute with each other and stabilize a given quantum state, meaning they leave the state invariant under their action. Formally, if $|\psi\rangle$ is a quantum state and S is an operator belonging to the stabilizer group, then $S|\psi\rangle = |\psi\rangle$.

The stabilizer formalism is useful because it allows for a compact representation of certain quantum error-correcting codes. For instance, in the three-qubit bit-flip code, the encoded logical state has the form $|\psi\rangle_L = \alpha|000\rangle + \beta|111\rangle$. The set of these error-free states, referred to as the code space, are described by the associated stabilizers ZZI , IZZ , and ZIZ . All states within the code space are simultaneous $+1$ eigenstates of these stabilizers.

Stabilizer place constraints on the permissible states of the code space. When an error affects a qubit, it alters the eigenvalues of the stabilizers. Measuring these stabilizers allows one to diagnose which qubit is in error and subsequently correct it, without disturbing the error-free logical states, which are eigenstates of the stabilizers.

For example, the error syndrome detection for the three-qubit bit-flip code shown in Fig. 12.12 employs two parity measurements: one between the first and second qubits, and another between the first and third qubits. These parity measurements correspond to measuring the stabilizers ZZI and ZIZ , respectively. Using only two out of the three stabilizers is sufficient because the product of any two stabilizers yields the third.

For the nine-qubit Shor code, the stabilizer group consists of operators that act on nine qubits, including:

$$\begin{aligned} & ZZIIIIII, \quad ZIZIIIII, \quad IIIIZZIIII, \quad IIIZIZIIII, \\ & IIIIIIZZI, \quad IIIIIIZIZ, \quad XXXXXXIII, \quad IIIIXXXXX. \end{aligned}$$

The above operators form the minimal set of independent stabilizers that uniquely define the code space, often referred to as stabilizer generators. Every state in the code space is a $+1$ eigenstate of each of these stabilizers. These stabilizer generators are specifically the operators that are measured in the error syndrome detection circuit, as illustrated in Fig. 12.15. Their measurements make it possible to identify

and correct both bit-flip (X) and phase-flip (Z) errors without disturbing the encoded quantum information.

Exercise 12.31 Explain why these are not stabilizers for the nine-qubit Shor code: $ZIIZIIIII$, $ZZZIIIIII$, $XXXIIIIII$, and $ZZZZZZIII$.

Exercise 12.32 Explain why these are stabilizers but not included as generators in the table above for the nine-qubit Shor code: $IZZIIIIII$, and $XXXIIIXXX$.

3 Code Distance

The code distance, often denoted as d , is a metric in quantum error correction that signifies the minimum number of qubit errors required to transform one valid encoded state (codeword) into another without detection by the code's syndrome measurements.

The three-qubit bit-flip code can correct a single bit-flip error, such as XII . It can also detect a double bit-flip errors like XXI , even though cannot correct it due to syndrome ambiguity with the single bit-flip error IIX . Therefore, the code distance for bit-flip errors is $d = 3$. On the other hand, the code is not designed to detect or correct phase-flip errors, rendering its code distance for phase-flip errors as $d = 1$.

In contrast, the nine-qubit Shor code is designed to correct for one arbitrary error on any of the nine qubits. It would take at least three such arbitrary errors to go undetected. Therefore, it has a code distance of $d = 3$ for both bit-flip and phase-flip errors.

12.4.5 *Further Exploration

The field of quantum error correction (QEC) is critical in advancing the development of large-scale, fault-tolerant quantum computers. Given that real-world quantum devices invariably encounter various types of noise and errors, a robust understanding and effective mitigation of these imperfections are fundamental. For those interested in delving deeper into error correction strategies and noise reduction techniques in quantum computing, the following topics and references are recommended for extended exploration:

1 Advancements in Efficient QEC

The development of quantum error-correcting codes has been a cornerstone in the advancement of quantum computing. The first such code, known as Shor's 9-qubit code, laid the foundation for subsequent innovations. Following this, Andrew Steane introduced a more efficient code, known as Steane's code, which could detect and correct any single-qubit error using only 7 physical qubits to encode 1 logical qubit[84]. Further progress was made by Raymond Laflamme and colleagues, who developed codes that required as few as 5 qubits[64].

Surface codes represent another significant class of quantum error-correcting codes. Characterized as topological quantum error-correcting codes, they are implemented on a two-dimensional lattice of qubits[34]. The error correction in surface codes leverages the lattice's topology. Although surface codes generally necessitate a larger

number of qubits than the likes of Steane's or Laflamme's codes, they are highly valued for their high fault tolerance and adaptability to two-dimensional grid layouts with nearest-neighbor qubit interactions.

Recent advancements in quantum error correction have seen the emergence of new codes, notably within the Low-Density Parity Check (LDPC) family, that promise a tenfold increase in efficiency compared to traditional methods[29]. These innovative LDPC codes achieve effective error correction with a reduced qubit requirement, marking a pivotal step towards the realization of scalable and practical quantum computing. This evolution reflects the ongoing progress in quantum error correction, moving from the original Shor's 9-qubit code to increasingly resource-efficient solutions.

An emerging strategy in quantum error correction is the concept of erasure conversion. This method transforms various quantum errors into erasure errors, whose locations are known, thereby simplifying their correction. Recognized for its potential to reduce error correction complexity and overhead, erasure conversion is seen as a key strategy for enhancing quantum computing efficiency. This approach could significantly streamline the error correction process, marking a crucial step towards scalable and practical quantum computing[49, 91].

2 General References on QEC

1. Quality, Speed, and Scale: Key Attributes of Near-Term Quantum Computers [88]: This paper emphasizes the three primary attributes to evaluate the performance of near-term quantum computers, guiding the assessment of quantum progress.
2. Foundations of Quantum Error Correction [34]: A comprehensive review tailored for computer scientists unfamiliar with quantum physics. It discusses the foundational principles of QEC, practicality of QEC codes, and challenges for NISQ computers.

3 Fault Tolerance Threshold

1. Comparison of Quantum Error-Correction Threshold for Exact and Approximate Errors [51]: This study uses classical simulations of stabilizer circuits to assess quantum code thresholds. It integrates Clifford and Pauli operators for improved noise approximation, focusing on the accuracy of Steane code and the Pauli twirling approximation.

4 Error Mitigation and Simulation

1. How to Simulate Quantum Measurement Without Computing Marginals [28]: A novel approach to quantum simulation that bypasses the need to compute marginals during quantum measurement.
2. Error Mitigation for Universal Gates on Encoded Qubits [73]: This work delves into advanced techniques for error mitigation when using universal gates on encoded qubits.
3. Mitigating Coherent Noise Using Pauli Conjugation [33]: A proposed method to mitigate coherent noise in quantum computations leveraging the Pauli conjugation technique.

4. Mitigating Depolarizing Noise on Quantum Computers [87]: This paper offers insights into noise-estimation circuits designed to counteract depolarizing noise, ensuring enhanced computational fidelity.

5 Holonomic and Topological Quantum Computation

Holonomic quantum computation (HQC) and topological quantum computation (TQC) are advanced approaches to quantum computing. Both leverage specific non-local properties of quantum systems for quantum operations, offering inherent resistance to certain types of errors.

The underlying principle of HQC involves exploiting the global properties (or geometry) of quantum systems, rather than their local properties (such as the quantum state). In HQC, computation is realized by cyclically moving the system within its parameter space (for instance, on the surface of the Bloch sphere). During this movement, the system acquires geometric phases (or Berry's phases) corresponding to the solid angles enclosed by its trajectory. As a result of relying on global properties, HQC displays a commendable resilience against local errors.

On the other hand, TQC derives its computational capability from the braiding of anyons. These anyons are exotic quasiparticles present in two-dimensional systems, distinguished from both fermions and bosons (see § 6.3.4). The topological nature of the quantum state, upon which the computation is built, ensures that local disturbances have minimal impact on the results. Such topological protection provides a potential pathway for fault-tolerant quantum computation.

As of early 2024, while HQC principles have seen experimental validation across several qubit systems, TQC remains more elusive. Although TQC boasts a robust theoretical foundation, its experimental realization presents significant challenges. Notably, the controlled observation and braiding of anyons remain at the forefront of quantum research.

For further exploration, readers are encouraged to consult the following references:

1. Enhancing Error Resilience Using Geometric Phase-Based Quantum Gates [83]: An introductory paper explaining the basics of geomtric phase and HQC, exploring the maximal error resilience achievable using quantum gates designed with geometric phase shifts, and their compatibility with other quantum error-correction techniques.
2. Berry's Phase in *Introduction to Quantum Mechanics* [3]: This book section delves into the concept of Berry's phase (geometric phase) and its relation to cyclic adiabatic evolution. The Berry phase is determined solely by the path in the parameter space, irrespective of the traversal rate along that path. The book exemplifies this using the physical system of an electron undergoing cyclotron motion, providing a comprehensive explanation in the dedicated section on Berry's phase.
3. Quantum Algorithm Design Exploiting Geometric and Holonomic Gates [97]: A recent paper offering an overview on the use of intrinsic geometric properties in quantum-mechanical state spaces for the realization of quantum logic gates. It also investigates the development of novel quantum algorithms that can take advantage of the error resilience and unique properties of geometric and holonomic quantum gates.

4. Harnessing Non-Abelian Phases for Robust Quantum Information Processing [85]: A comprehensive overview of leveraging non-Abelian quantum phases in topological quantum computation, emphasizing their nonlocal nature that renders quantum information resilient against environmental interactions and protocol imperfections. The review further highlights various solid-state systems, such as those hosting Majorana fermions and non-Abelian quantum Hall states, as potential platforms for implementing these concepts.
5. Non-Abelian Anyons: A Cornerstone of Fault-Tolerant Quantum Computation [70]: An in-depth exploration of the significance of non-Abelian anyons in the realm of topological quantum computation. The article discusses the unique braiding statistics of these anyons and how they enable fault-tolerant quantum gate operations. Emphasis is placed on the inherent resilience imparted by the nonlocal encoding of quasiparticle states, detailing potential architectures and experiments for realizing a topological quantum computer.

12.5 Summary and Conclusions

Essentials of Quantum Error Correction

In this chapter, our emphasis was on elucidating the nuances of quantum error correction, a domain that has emerged as vital for the progression and practicality of quantum computing. Given the inherent sensitivity of quantum systems to noise and their potential to lose coherence, understanding errors, their causes, and methods to rectify them is paramount.

We began our journey by emphasizing the importance of preliminary concepts, shedding light on the challenges posed by noise and decoherence. By drawing distinctions between decoherence, noise, and errors, we underscored the unique features and manifestations of each, leading to a clearer understanding of the quantum noise problem.

Density Operators and Mixed States

Transitioning from state vectors, we introduced the mathematical framework of density operators, a powerful tool to capture the subtleties of mixed quantum states. The relevance of density operators in modeling diverse quantum operations, especially in the context of noise and decoherence, was highlighted.

We navigated through their fundamental properties, delving into the computation of expected values, the representation of qubit states, and their interpretation on the Bloch sphere. This not only expanded our understanding of quantum states but also provided a lens to appreciate the richness of quantum mechanics.

Error Mechanisms and Error Correction

Our discourse then veered towards the core focus: the diverse error mechanisms in quantum computing. We identified and elaborated on coherent errors, characterized by their systematic nature, and incoherent errors, which arise due to stochastic interactions with the environment. Through examples such as bit-flip, phase-flip, and depolarizing errors, we accentuated the varied manifestations of these errors and the challenges they pose.

To combat these challenges, we embarked on an exploration of quantum error

correction codes. Here, the emphasis was on introducing and dissecting the principles and techniques that serve as our arsenal against quantum errors. We delved into the Shor codes, highlighting their significance in the realm of quantum error correction, and underscored the importance of parity measurements, stabilizers, and code distance in this context.

Bridging Theory and Practice

The overarching narrative of this chapter bridges the theoretical underpinnings of quantum mechanics with the pragmatic challenges faced in quantum computing. By intertwining the mathematical formalism of density operators with real-world error mechanisms, we have endeavored to prepare the reader for both theoretical research and hands-on experimentation in quantum error correction.

The journey through this chapter offers a holistic perspective on the challenges and solutions in the domain of quantum error correction, laying the groundwork for the subsequent discussions and further explorations in this dynamic field.

Upcoming Topics

The upcoming chapter provides an exploration of quantum information theory. It begins by contrasting quantum probability with classical probability, highlighting unique quantum aspects like entanglement. The chapter then discusses quantum entropy and information, comparing these with classical concepts. Finally, it delves into advanced topics like the quantum data processing inequality and the Holevo theorem, emphasizing their theoretical and practical importance in quantum communications.

Problem Set 12

12.1 Derive the density operator ρ_{\perp} for the qubit state orthogonal to $|\psi\rangle$: $|\psi_{\perp}\rangle = \beta^*|0\rangle - \alpha^*|1\rangle$. Show that ρ and ρ_{\perp} are orthogonal to each other, i.e., $\text{tr}(\rho^{\dagger}\rho_{\perp}) = 0$. In fact, $\rho^{\dagger}\rho_{\perp} = 0$.

12.2 Derive the density operator for the general qubit state $|\psi\rangle = \cos \frac{\theta}{2}|0\rangle + \sin \frac{\theta}{2}e^{i\phi}|1\rangle$.

12.3 Alice and Bob share pairs of qubits, with half of the pairs in the Bell state $\frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$ and the remaining half in the state $\rho_A \otimes \rho_B$, where $\rho_A = \frac{1}{4} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $\rho_B = |0\rangle\langle 0|$.

Determine the density operator for the composite system of Alice and Bob's shared qubits.

12.4 Represent the following quantum states on (or in) the Bloch sphere:

- 20% in $|0\rangle$ and 80% in $|1\rangle$
- 50% in $|+\rangle$ and 50% in $|0\rangle$
- 25% of $|0\rangle$, 25% of $|1\rangle$, and 50% of $|+\rangle$

(d) 50% in $|+\rangle$ and 50% in $|-\rangle$

12.5 The state of a quantum system is such that it is 50% in the state $|+\rangle$ and 50% in the state $|0\rangle$. A Hadamard gate is applied to the system. Find the new density operator for the system.

12.6 How does the Hadamard gate H transform a quantum state represented by its Bloch vector $\vec{r} = (x, y, z)$?

12.7 Given the initial state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, what is the density operator of the ensemble of measurement outcomes upon measurements in the $\{|0\rangle, |1\rangle\}$ basis?

12.8 The state of a system, initially 50% in the state $|+\rangle$ and 50% in the state $|0\rangle$, is measured in the $\{|0\rangle, |1\rangle\}$ basis. What is the density operator of the post-measurement ensemble?

12.9 Consider the GHZ state, which is defined for three qubits as $|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$.

- (a) Calculate the reduced density operator of the first two qubits after the third qubit has been traced out (i.e., its degrees of freedom have been disregarded).
- (b) Calculate the reduced density operator of the third qubit after the first two qubits are traced out.

12.10 Consider the three-qubit W state, defined as $|\text{W}\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$, known for its characteristic partial entanglement among qubits without any pair being maximally entangled. Your task involves:

- (a) Constructing the density operator ρ_{W} for the W state.
- (b) Calculating the reduced density matrices for each pair of qubits by performing the appropriate partial traces on ρ_{W} .
- (c) Confirming the partial entanglement between each pair of qubits through analysis of their reduced density matrices.

12.11 Investigate the two-qubit cross-talk error caused by unintended ZX coupling, which is represented by the Hamiltonian $H_{ZX} = -\frac{\hbar\omega}{2}(Z \otimes X)$.

- (a) The undesired ZZ coupling is known to result in a phase walk-off in the computational basis, as discussed in § 12.3.1. Describe the specific type of error that arises from unintended ZX coupling.
- (b) Derive the equations that quantify the cross-talk errors due to ZX coupling in a manner analogous to the errors described by Eqs. 12.86 and 12.87.

Hint: Consult § 6.5.4 for a discussion on ZX coupling and its implications in the context of CNOT gate implementation.

12.12 Consider a qubit initially in the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. Investigate the impact of the following incoherent errors on the qubit, expressing the resultant state as a density operator and visualizing the changes on the Bloch sphere. Detailed discussions of these errors can be found in § 12.3.2.

- (a) Bit-flip error
- (b) Phase-flip error

- (c) Depolarizing error
- (d) Dephasing error
- (e) Amplitude damping error
- (f) Pauli channel error

12.13 Consider a two-qubit system initially in the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Analyze the state of the system expressed as a density operator after it undergoes the following incoherent errors. For detailed discussions on these errors, refer to § 12.3.2.

- (a) Two-qubit bit-flip error
- (b) Two-qubit dephasing error
- (c) Two-qubit Pauli channel

12.14 Unlike classical computers, which largely rely on a homogeneous technology base, quantum computers are being developed using a variety of hardware platforms, each advancing rapidly. For this exercise, select three distinct quantum computing hardware platforms (e.g., superconducting qubits, trapped ion qubits, neutral atom qubits, photonic systems, etc.). For each selected platform, conduct a detailed investigation into the following:

- The specific aspects and applications where it excels.
- The challenges or limitations it faces.

Quantify these strengths and weaknesses using the performance metric parameters outlined in § 12.3.4, such as coherence time, qubit connectivity, gate fidelity, error rates, quantum volume, and scalability, among others.

Prepare a presentation on your findings, ensuring to:

- Clearly define each hardware platform and its operational principles.
- Discuss and quantify the performance metrics for each platform, providing a comparative analysis.
- Highlight the implications of these metrics for future developments and applications of quantum computing.

Consider utilizing reputable scientific journals, technology reports, and official documentation from research institutions as sources for your research to ensure accuracy and relevancy.

This exercise aims to deepen your understanding of the diverse landscape of quantum computing technologies and their potential impact on the field's evolution.

12.15 Construct an alternative version of the Shor three-qubit bit-flip code, as detailed in Fig. 12.12. In this variant, modify the pairs of qubits used for parity measurement. Instead of employing the qubit pairs (q_1, q_2) and (q_1, q_3) , use the pairs (q_1, q_2) and (q_2, q_3) .

12.16 In Fig. 12.13, the error syndrome measurement for the three-qubit phase-flip code utilizes two sets of Hadamard gates on the data qubits. Design a circuit

that performs the same syndrome measurement function, but without using these Hadamard gates.

Hint: Refer to the equivalent gate sequences discussed in § 7.4.

- 12.17** A logical Pauli Y gate, denoted as Y_L , transforms $\alpha|0_L\rangle + \beta|1_L\rangle$ into $-i\beta|0_L\rangle + i\alpha|1_L\rangle$. Design four distinct versions of the logical Y_L gate for the nine-qubit Shor code.

Hint: Recall that $Y = iXZ$. Refer to Exercise 12.28 for examples of X_L and Z_L implementations.

- 12.18** The Shor code shown in Fig. 12.14 begins with a phase-flip block followed by three bit-flip blocks. Investigate the consequences of starting with a bit-flip block and then following with three phase-flip blocks.

13. Fundamentals of Quantum Information

Contents

13.1	Quantum Probability Essentials	399
13.1.1	Distinctive Features of Quantum Probability	399
13.1.2	Sample Space, Events, and Probability Distribution	400
13.1.3	Random Variables and Expected Values	402
13.1.4	Multiple Random Variables	404
13.1.5	Random Processes	412
13.1.6	* Example: One-Dimensional Quantum Walk	413
13.2	Quantum Entropy and Information	417
13.2.1	Classical Shannon Entropy	418
13.2.2	* Quantum von Neumann Entropy	420
13.2.3	* Classical Joint Entropy and Mutual Information	421
13.2.4	* Quantum Joint Entropy and Mutual Information	425
13.3	* Core Theorems in Quantum Information	428
13.3.1	* The Data Processing Inequality	429
13.3.2	* The Holevo Bound and Channel Capacity	431
13.3.3	* Subadditivity and Strong Subadditivity	434
13.4	Further Exploration	435
13.5	Summary and Conclusions	435
	Problem Set 13	436

The field of information theory underwent a significant shift with the introduction of quantum mechanics. The principles that govern the world of atoms and photons provided a new perspective for reimagining the foundation of information. This chapter begins an explorative journey from the familiar territory of classical information to the mysterious and less intuitive realm of quantum information.

Classical information theory, rooted in the seminal work of Claude Shannon, laid the foundation for understanding the transmission, processing, and encoding of information within a probabilistic framework. Central to this theory is the concept of Shannon entropy, which quantifies the uncertainty or information content

in a classical system. Classical systems, governed by deterministic laws, utilize probabilities to address the incomplete knowledge about the system. Despite their deterministic nature, probabilistic methods in classical information theory are crucial for managing uncertainties inherent in information transmission and processing.

In contrast, quantum information theory is deeply intertwined with the inherent probabilistic nature of quantum mechanics, a fundamental departure from classical theories. In quantum mechanics, probabilities do not merely represent a lack of knowledge but are an intrinsic aspect of quantum states. This intrinsic probabilistic nature is encapsulated by the Born rule, which governs the probabilities of various outcomes when measurements are made on quantum states. Consequently, quantum information theory offers a richer framework for encoding and processing information, taking into account the phenomena of superposition and entanglement. Analogous to Shannon entropy in classical systems, von Neumann entropy extends these concepts to the quantum realm, accommodating the unique properties of quantum states and their non-classical correlations.

The significance of quantum probability and entropy transcends the mere extension of classical concepts into the quantum domain. They are the cornerstone for understanding quantum coherence and correlations, which have no classical counterpart. Quantum entropy not only measures information but also captures the degree of entanglement—a unique resource for quantum computing and communication. The interplay between quantum probability, entropy, and information fosters new potential for advancements that include the enhanced efficiency of quantum algorithms, the establishment of secure communication channels, and the development of quantum error correction techniques.

In exploring the fundamentals of quantum information, we lay the groundwork that will enable researchers and practitioners to actively engage in and further the quantum revolution in computation, communication, and measurement. This foundational knowledge is essential for those who aspire to drive the forthcoming era of technological innovations in quantum information science.

Prerequisite Review: Mixed States and Density Operators

Mixed states and density operators are detailed in § 12.2. Here is a brief summary of the concepts and formulas essential for this chapter.

- A mixed quantum state is a statistical ensemble of multiple potential pure states. It encapsulates both quantum uncertainty and classical uncertainty. A mixed qudit state with probability p_i of $|\psi_i\rangle \in \mathbb{C}^d$, $i = 1, \dots, n$, can be described by a density operator:

$$\rho = \sum_{i=1}^n p_i |\psi_i\rangle\langle\psi_i|. \quad (13.1)$$

- Properties of density operators:
 - Hermitian: $\rho^\dagger = \rho$
 - Positive Semidefinite: $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathbb{C}^d$
 - Unit Trace: $\text{tr}(\rho) = \sum_{i=1}^d \rho_{ii} = 1$
- A density operator has a diagonal form in its eigenbasis, corresponding to the canonical states of the quantum system:

$$\rho = \sum_{i=1}^d \lambda_i |\lambda_i\rangle\langle\lambda_i|, \quad (13.2)$$

with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$.

- Applying a unitary transformation $U \in \mathbb{C}^{d \times d}$ transforms ρ as

$$\rho \rightarrow U\rho U^\dagger. \quad (13.3)$$

- Measuring ρ in orthonormal basis $\{|\phi_i\rangle\}$ yields outcome $|\phi_i\rangle$ with probability:

$$p_i = \langle\phi_i| \rho |\phi_i\rangle = \text{tr}(\rho |\phi_i\rangle\langle\phi_i|). \quad (13.4)$$

Following the measurement, the ensemble of measurement outcomes is given by

$$\rho_M = \sum_{i=1}^d p_i |\phi_i\rangle\langle\phi_i|, \quad (13.5)$$

which is diagonal in the measurement basis. This represents the equivalence of state collapse in the context of mixed states.

In the more general POVM measurement framework, a measurement outcome corresponds to a POVM operator E_i , and the probability p_i of observing outcome i is given by:

$$p_i = \text{tr}(\rho E_i). \quad (13.6)$$

- For a composite system consisting of two independent subsystems, A described by ρ_A and B by ρ_B , the combined density operator is given by

$$\rho_{AB} = \rho_A \otimes \rho_B. \quad (13.7)$$

- Partial trace of a density operator provides a way to focus on a subset of a larger system by summing over (and effectively ignoring) the rest of the system. The partial trace over subsystem A, or the reduced density operator of subsystem B, is given by

$$\rho_A = \text{tr}_B(\rho_{AB}) = \sum_i \langle i|_B \rho_{AB} |i\rangle_B, \quad (13.8)$$

where $\{|i\rangle_B\}$ is any orthonormal basis for the Hilbert space of subsystem B.

Density Operator: Pillar of Quantum Probability and Information

The density operator plays key roles in quantum probability and information, as a versatile and foundational construct in quantum mechanics, underpinning concepts such as von Neumann entropy and quantum information.

- The density operator ρ , being Hermitian, can be diagonalized to reveal a probability distribution over its constituent eigenstates. Each eigenvalue λ_i of ρ corresponds to the probability of finding the system in the associated eigenstate $|\lambda_i\rangle$. This structure allows ρ to represent both pure and mixed states in a unified framework.

- In quantum communication, the density operator can represent mixed states arising from noise and environmental interactions, essential for quantum error correction and channel capacity.

When a state ρ is measured in an orthonormal basis $\{|\phi_i\rangle\}$, the resulting post-measurement ensemble is captured by a new density operator ρ_M . Diagonal elements of ρ_M , given by $\langle\phi_i|\rho|\phi_i\rangle$, form a probability distribution over the measurement basis states, reflecting the outcome probabilities. (See discussion around Eq. 12.48.)

- Quantum coherence, encapsulated in the off-diagonal elements of ρ in a given basis, is fundamental to quantum computation and information processing. It represents the superposition of states and is a key resource for quantum algorithms, enabling quantum parallelism and interference.
- Purity and entropy, critical in quantum thermodynamics and information theory, are derived from the density operator's properties. Purity, indicating the degree of mixedness of a state, is given by $\text{tr}(\rho^2)$, while von Neumann entropy, measuring the informational content, is $-\text{tr}(\rho \log \rho)$.

13.1 Quantum Probability Essentials

Quantum information science is deeply rooted in the principles of probability theory. In this section, we examine the fundamental components of probability theory as they apply to quantum information science. Beginning with a discussion of each concept within the classical framework, we then establish parallels with their quantum mechanical counterparts. This comparative analysis enhances our understanding of the inherently probabilistic nature of quantum systems.

We will use the following convention for symbols unless stated otherwise:

Symbol	Meaning
X, Y	Random variables
x, y	Values of random variables X and Y , respectively
A, B, C	Subsystems or random events
p	Probability or probability distribution
$p(x)$	Probability that X takes the value x
$p(X)$	Probability distribution of X
$H(p), H(X)$	Shannon entropy of a prob. distribution p or random variable X
$S(\rho), S(X)$	von Neumann entropy of density operator ρ or random var X
$E[X]$	classical expected value (statistical average) of X
$\langle H \rangle$	quantum expected value (statistical average) of observable H
$\sigma_x, \sigma_y, \sigma_z$	Pauli operators

13.1.1 Distinctive Features of Quantum Probability

While quantum probability encompasses the general structure of classical probability, it introduces several key features that set it apart, fundamentally altering our understanding of probabilistic phenomena.

1 Complex-Valued Probability Amplitudes

In classical probability, probabilities are real numbers between 0 and 1. Quantum mechanics, however, uses complex-valued probability amplitudes. These amplitudes, when squared, give the probabilities of various outcomes. The use of complex numbers allows for the phenomenon of quantum superposition, where a quantum state can simultaneously exist in multiple states. This superposition is a cornerstone of quantum mechanics, leading to interference effects that have no counterpart in classical probability.

2 Entanglement

Entanglement introduces correlations between quantum systems that are non-existent in classical systems. When two particles are entangled, their quantum states become so interlinked that the state of one particle correlates with the state of the other, regardless of the distance separating them. This leads to joint probability distributions that cannot be decomposed into independent probabilities of each entangled component. This nonlocal property of entangled systems profoundly affects the interpretation and calculation of probabilities in quantum systems.

3 Non-commutativity of Quantum Observables

Another fundamental aspect of quantum probability is the non-commutative nature of quantum observables. When two operators do not commute, the precise sequence of measurements affects the observed probabilities. This is exemplified by the Heisenberg uncertainty principle (see § 1.6), which states that certain pairs of observables, like rectilinear and circular polarizations, cannot be simultaneously measured with arbitrary precision. The non-commutativity of observables can lead to complex probability distributions and correlations that cannot be replicated in classical probability theory.

13.1.2 Sample Space, Events, and Probability Distribution

1 Classical Probability

Random Experiments

In classical probability, a random experiment or trial is defined as a process with a definitive set of possible outcomes, which cannot be precisely predicted in advance. Such an experiment can be replicated under consistent conditions, with each repetition being independent of the others.

Sample Space

The sample space, denoted $\Omega = \{s_1, s_2, \dots, s_n\}$, encompasses all conceivable outcomes of an experiment, with each s_i representing a unique outcome. This space can be finite or countably infinite for discrete cases and must fulfill two criteria:

1. Exhaustiveness: It should encompass all potential outcomes.
2. Mutual Exclusivity: No two outcomes should simultaneously occur in a single trial.

Events

An event, denoted as E , is any subset of the sample space, $E \subseteq \Omega$. An event E occurs if the outcome of the random experiment is contained within E . An event

may be an elementary event (consisting of a single outcome) or a compound event (comprising multiple outcomes).

■ **Example 13.1** The roll of a six-sided die constitutes a random experiment, with the sample space represented by $\Omega = \{1, 2, 3, 4, 5, 6\}$.

The event of rolling an even number on the die is $E = \{2, 4, 6\}$. The event of rolling an odd number on the die is $O = \{1, 3, 5\}$. E and O are exhaustive and mutually exclusive. ■

Probability Distribution

A probability distribution assigns a probability to each experiment outcome. These assignments adhere to:

1. Non-Negativity: Probabilities are always non-negative.
2. Normalization: The sum of all outcome probabilities equals 1.
3. Additivity: The probability of the union of two distinct events is the sum of their individual probabilities.

■ **Example 13.2** For a fair die, the probability distribution over Ω is uniform:

$$p_i \equiv p(\{i\}) = \frac{1}{6} \quad \text{for each } i \in \Omega. \quad (13.9)$$

Thus, the probability of event E (rolling an even number) is:

$$p(E) = p(\{2\}) + p(\{4\}) + p(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}. \quad (13.10)$$

■

2 Quantum Generalization

Sample Space: Spectrum of an Observable

In quantum mechanics, the classical concept of a sample space is reinterpreted as the spectrum of an observable, represented by a Hermitian operator H . The eigenvalues $\{\lambda_i\}$ of this operator represent the possible results we can obtain when measuring the observable. Each eigenvalue has a corresponding eigenvector (or a subspace when there's degeneracy), which represents the quantum state the system will be in if we obtain that specific measurement outcome (see § 3.4).

More generally, a quantum sample space can be represented by an orthonormal basis in the Hilbert space of the system, where measurements are conducted.

Events: Quantum Measurements

In the quantum realm, events are paralleled by measurements. For an observable H , distinct eigenvalues correspond to orthogonal eigenstates $\{|\lambda_i\rangle\}$, analogous to classical mutually exclusive events. These eigenstates form a complete basis set in the Hilbert space, akin to a complete set of events.

More generally, quantum measurements can be described by POVM (Positive Operator-Valued Measure) operators $\{E_i\}$. While some scenarios allow E_i to correspond to projective measurements like $|\lambda_i\rangle\langle\lambda_i|$, associated with the eigenstates of observables, POVMs offer a more generalized framework and can represent a

broader set of outcomes. For example, two POVM operators could represent the aggregate events of rolling an even or odd number on a die.

Probability Distribution: The Born Rule

The Born rule dictates the probabilities of measurement outcomes in quantum mechanics, forming the basis for quantum probability distribution. The probabilistic behavior of a quantum system is fully characterized by its state vector $|\psi\rangle$ for pure states or density operator ρ for mixed states. The state vector $|\psi\rangle$ is associated with the inherent quantum randomness of measurements, while ρ accounts for both quantum and classical randomness due to statistical mixtures of states.

The probability p_i of observing a particular outcome λ_i of an observable is given by:

$$p_i = \begin{cases} |\langle \lambda_i | \psi \rangle|^2 = \langle \psi | \lambda_i \rangle \langle \lambda_i | \psi \rangle & \text{for a pure state,} \\ \langle \lambda_i | \rho | \lambda_i \rangle = \text{tr}(\rho | \lambda_i \rangle \langle \lambda_i |) & \text{for a mixed state.} \end{cases} \quad (13.11)$$

For POVM measurements, outcome probabilities for each E_i are:

$$p_i = \begin{cases} \langle \psi | E_i | \psi \rangle & \text{for a pure state,} \\ \text{tr}(\rho E_i) & \text{for a mixed state.} \end{cases} \quad (13.12)$$

13.1.3 Random Variables and Expected Values

1 Classical Probability

A random variable is a function that assigns a numerical value (or other mathematical constructs) to each outcome in the sample space of a random experiment. For example, a random variable X representing the result of rolling a die conventionally assumes values in the set $\{1, 2, 3, 4, 5, 6\}$, but it may also assume values in some other set, e.g., $\{-3, -2, -1, 1, 2, 3\}$. The assignment may even be many-to-one, e.g., $\{o, e, o, e, o, e\}$.

As introduced above, a probability distribution assigns a probability to each outcome in a sample space. While a probability distribution originates from the concept of a sample space, it can also be defined in the context of random variables. In the latter case, a probability distribution describes how probabilities are distributed over the different possible values that the random variable can take. In the above example of rolling a die, the probability distribution can be expressed as $p(X = x) = \frac{1}{6}$ for $x = 1, 2, \dots, 6$.



When a random variable employs a many-to-one assignment to map outcomes in the sample space to numerical values, it effectively aggregates those outcomes under the assigned values. This aggregation may result in a probability distribution for the random variable that differs from the original distribution of outcomes in the sample space.

For example, let X represent whether the outcome of rolling a die is even or odd. It takes the value e for outcomes $\{2, 4, 6\}$ and o for $\{1, 3, 5\}$. Now, $p(X = e) = p(X = o) = \frac{1}{2}$.

Events can also be described in terms of random variables. For example,

$E = \{X \leq 3\}$ is an event representing the outcome of rolling a die taking on a value at most 3.

The expected value (or expectation value) quantifies the statistical average of a random variable, weighted by its probability distribution, indicating the central tendency of the distribution. For instance, the expected value of a fair die's roll is

$$E[X] = \sum_{x=1}^6 x \cdot p(X = x) = 3.5. \quad (13.13)$$

2 Quantum Generalization

An observable in quantum mechanics can be considered a generalization of a classical random variable because it describes the possible outcomes of a measurement (a random process) and their associated probabilities. However, unlike classical random variables, which directly map outcomes to real numbers, the outcomes of measuring an observable are determined by the eigenvalues of its operator, and the probability of each outcome is given by the state of the system before measurement.

Much like in the classical case, the expected value of an observable tells us about the average result we'd see if we measured it many times, given the quantum state of the system. The expected value of an observable H is given by:

$$\langle H \rangle = \begin{cases} \langle \psi | H | \psi \rangle & \text{for a pure state,} \\ \text{tr}(\rho H) & \text{for a mixed state.} \end{cases} \quad (13.14)$$

The variance of an observable H , which is equal to the square of the standard deviation and quantifies the spread of measurement outcomes (see § 3.4.6), is calculated as:

$$(\Delta H)^2 = \langle H^2 \rangle - \langle H \rangle^2 = \begin{cases} \langle \psi | H^2 | \psi \rangle - \langle \psi | H | \psi \rangle^2 & \text{for a pure state,} \\ \text{tr}(\rho H^2) - (\text{tr}(\rho H))^2 & \text{for a mixed state.} \end{cases} \quad (13.15)$$

Exercise 13.1 Consider a two-qubit system where 50% of the system is in the uniform mixed state described by the density operator $\frac{1}{4}I$, with I being the 4×4 identity operator, and 50% of the system is in the Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. The system is measured with the observable $H = Z \otimes Z$, where $Z \equiv \sigma_Z$ denotes the Pauli-Z operator. Calculate:

- (a) The density operator of the system, ρ .
- (b) The eigenvalues and eigenstates of the observable H .
- (c) The probability of obtaining each eigenvalue of H as the measurement outcome.
- (d) The expected value of the measurement, $\langle H \rangle$.
- (e) The variance of the measurement outcomes, $(\Delta H)^2$.

Key Takeaways

In quantum probability, quantum observables fulfill the role of random variables. The density operator encapsulates both classical and quantum randomness of a system. Measurements correspond to probabilistic events and are made with respect to the eigenbasis of an observable, serving as the sample space. The probability distribution is governed by the Born rule.

13.1.4 Multiple Random Variables

1 Classical Probability

Analyzing multiple random variables involves understanding their interactions and the consequent impact on their probabilities. Although we focus on two random variables, for simplicity, these principles extend to more variables.

Probabilities of Two Random Variables

Consider two random variables, X and Y , defined over the same sample space Ω . We examine specific events associated with these variables, represented by the outcomes x for X and y for Y . The probabilities linked to these events are outlined as follows:

- Joint Probability: The probability of x and y occurring together is denoted as $p(X = x, Y = y)$, $p(x \cap y)$, or simply $p(x, y)$, where $p(x, y) = p(y, x)$. The joint probability distribution, denoted as $p(X, Y)$, represents $p(x, y)$ for all possible values of x and y .
- Conditional Probability: This is the probability of event x given event y , denoted as $p(x|y)$. Note that $p(x|y)$ may not equal $p(y|x)$. We also extend this concept to the conditional probability distributions, $p(X|Y)$ and $p(Y|X)$.
- Marginal Probability: The probability of an event associated with one random variable, irrespective of the other variable, is calculated by summing over all outcomes of the other variable:

$$p(x) = \sum_{y \in \Omega} p(x, Y = y). \quad (13.16)$$

This process, known as “marginalization,” can also be expressed for the probability distribution:

$$p(X) = \sum_{y \in \Omega} p(X, Y = y). \quad (13.17)$$

- Probability of the Union: The probability that at least one of the events x or y occurs is denoted as $p(x \cup y)$.

Relationships Between Two Events

- Independent Events: The occurrence of one event does not affect the other. For example, the outcomes of two rolls of a die are independent. The probabilities for independent events satisfy:

$$p(x, y) = p(x) \cdot p(y), \quad p(x|y) = p(x). \quad (13.18)$$

- Dependent Events: One event influences the occurrence of the other, such as drawing cards from a deck without replacement. Their probabilities are related by the chain rule:

$$p(x, y) = p(y, x) = p(x|y) \cdot p(y) = p(y|x) \cdot p(x). \quad (13.19)$$

This leads to Bayes' Theorem:

$$p(x|y) = \frac{p(y|x) \cdot p(x)}{p(y)}. \quad (13.20)$$

- Mutually Exclusive Events: These events cannot occur simultaneously, implying:

$$p(x, y) = 0, \quad p(x \cup y) = p(x) + p(y). \quad (13.21)$$

- Overlapping Events: These events can occur together ($p(x \cap y) \neq 0$) and satisfy the inclusion-exclusion principle:

$$p(x \cup y) = p(x) + p(y) - p(x, y). \quad (13.22)$$

This is visually represented in a Venn diagram as shown in Fig. 13.1.

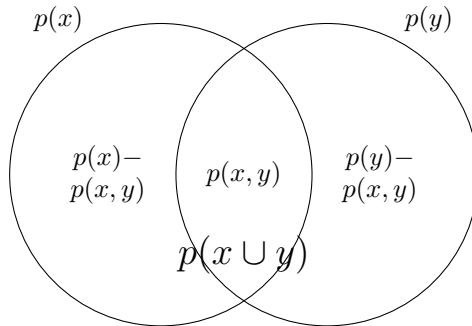


Figure 13.1: Venn Diagram for Probability Relationship

Normalization Conditions

The sum of the probabilities for all possible outcomes of a random variable X equals 1, indicating that some outcome in the sample space will occur with certainty. This normalization condition extends to joint and conditional probabilities:

$$\sum_{x \in \Omega} p(x) = 1, \quad (13.23a)$$

$$\sum_{x \in \Omega} \sum_{y \in \Omega} p(x, y) = 1, \quad (13.23b)$$

$$\sum_{x \in \Omega} p(x|y) = 1. \quad (13.23c)$$

The normalization for conditional probability (Eq. 13.23c) is valid for any given $y \in \Omega$ where $p(y) > 0$.

■ **Example 13.3 — Classical Joint Distribution and Marginal Probability.** Consider two random variables, X and Y , each taking values in $\Omega = \{1, 2, \dots, 6\}$ for fair dice.

Given two events: x is rolling a 1 on die one, and y is rolling a 3 on die two. The joint probability $p(x, y)$, the probability of both x and y occurring together, is given by the product of the probabilities of x and y , as the dice rolls are independent:

$$p(x, y) = p(x) \cdot p(y) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}. \quad (13.24)$$

The probability of the union of events x and y , $p(x \cup y)$, representing either event occurring, is given by:

$$p(x \cup y) = p(x) + p(y) - p(x, y) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}. \quad (13.25)$$

Now, we impose the condition that $X + Y$ is even. In this case, X and Y are no longer independent. The joint probability distribution $p(X, Y)$ becomes uniform over the subset of $\Omega \times \Omega$ where the sum of the dice is even, including pairs such as $\{(1, 1), (1, 3), (2, 2), (2, 4), \dots\}$, each with a probability of $\frac{1}{18}$. However, $p(X, Y)$ is not uniform over $\Omega \times \Omega$, as $p(X, Y) = 0$ if $X + Y$ is odd.

If Alice “holds” X and Bob “holds” Y , using the marginalization formula for Alice’s marginal distribution

$$p(X) = \sum_{y \in \Omega} p(X, Y = y), \quad (13.26)$$

we discover that $p(X)$ is still uniform over Ω , each outcome having a probability of $\frac{1}{6}$. Similarly, for Bob’s $p(Y)$.

This shows that if X and Y are dependent, the marginal distributions can be uniform over Ω , despite their joint distribution $p(X, Y)$ not being uniform over $\Omega \times \Omega$. We will see a similar phenomenon with a quantum EPR pair in Example 13.4. ■

2 Quantum Generalization

The concepts of classical multi-variable probability are extended to the quantum realm, introducing additional complexities. This complexity arises primarily from the non-commutative nature of quantum observables, where, unlike their classical counterparts, the order of measurement affects the outcome. Furthermore, quantum states of composite systems often exhibit entanglement, a phenomenon where the state of each component cannot be described independently of the others. Additionally, measurements on these systems can be either local, affecting individual subsystems, or joint, involving multiple subsystems simultaneously (see § 6.2.2). We will use examples to build connections with classical probabilities and illustrate these complexities.

Exercise 13.2 Prerequisite Review: To understand quantum probability and information concerning multi-qubit states, readers are highly recommended to review Exercise 6.6, replicated below for convenience.

Consider three qubits (A , B , and C) in the state

$$|\psi\rangle = \sum_{i,j,k \in \{0,1\}} c_{ijk} |ijk\rangle.$$

- (a) Find the probability of measuring the third qubit (C) with an outcome $|0\rangle$.
- (b) After qubit C is measured and collapsed to $|0\rangle$, what is the joint state of qubits A and B ? What is the state of the three-qubit system?
- (c) Suppose a Bell measurement is performed on qubits A and B with an outcome $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. What is the probability of this outcome? What is the post-measurement state of the three-qubit system?

Commuting vs. Non-Commuting Observables

A real-valued quantum random variable corresponds to an observable X , which is a Hermitian operator. The eigenvalues of X represent its possible outcomes.

As detailed in § 1.6, if two observables X and Y belonging to the same quantum system do not commute, i.e., $[X, Y] \neq 0$, there is generally no state in which both X and Y have definite values simultaneously; they do not share an eigenvector. In practical terms, two such variables are mutually uncertain and are not simultaneously measurable.

For instance, consider the non-commuting Pauli operators σ_X and σ_Z for a single qubit in state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. The measurement of one observable affects the system in a way that the subsequent measurement of the other becomes completely uncertain. (See exercise below.)

Thus, in quantum mechanics, the concept of joint and conditional probabilities for non-commuting observables does not apply in the same way it does for commuting observables or in classical probability theory.

Exercise 13.3 Consider the non-commuting Pauli operators σ_X and σ_Z for a single qubit in state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Compare two measurement sequences: measuring σ_X and then σ_Z on the qubit, and measuring σ_Z and then σ_X . Describe how the first measurement affects the state of the system and thereby the outcome of the second measurement. Discuss the meaning of $\langle \sigma_X \sigma_Z \rangle$ in this context.

Exercise 13.4 Consider two commuting Hermitian operators U and V on a single-qubit system: $U = V = \sigma_Z$, where σ_Z is the Pauli-Z operator. Verify if $\langle UV \rangle = \langle U \rangle \langle V \rangle$ for the following states:

- (a) $|\psi\rangle = |0\rangle$.
- (b) $|\psi\rangle = |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.
- (c) $\rho = \frac{1}{2}I$.

Marginal Probability Through Partial Trace

The reduced density operator calculated through the partial trace, discussed extensively in § 12.2.8, is the quantum counterpart of classical marginal probability. Both methodologies enable us to focus on a component within a composite system

by summing over and thereby effectively disregarding the other components of the system.

■ **Example 13.4 — Marginal Probabilities in an EPR Pair.** Let's re-examine the archetypal case where Alice and Bob share an EPR pair, $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. (See Example 12.10 for more details.)

To calculate the marginal probability, we start with the density operator for $|\Phi^+\rangle$:

$$\rho_{AB} = |\Phi^+\rangle\langle\Phi^+|. \quad (13.27)$$

The reduced density operator for Alice, obtained by tracing ρ_{AB} over B , captures the marginal probability of her qubit:

$$\rho_A = \text{tr}_B(\rho_{AB}) = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (13.28)$$

This is a maximally mixed state, with a uniform probability distribution over $|0\rangle$ and $|1\rangle$, in fact in all measurement bases.

Bob's reduced density operator is identical to Alice's, also showing a uniform probability distribution over $|0\rangle$ and $|1\rangle$. ■

Probabilities in a Product State

Let's now examine local measurements of a composite state of two qubit A and B . In this case, two observables, X and Y , pertaining to qubit A and B , respectively, commute regardless of their commutation properties within their respective systems. This is because the operations act on different Hilbert spaces and do not interfere with each other. (See § 7.4.1.)

If the state is a product state, measurements on one qubit are independent of measurements on the other, similar to the classical independent events.

■ **Example 13.5 — Joint and Conditional Probabilities in a Product State.** Consider two qubits A and B in the product state $|\psi\rangle = |++\rangle$, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.

(1) Measuring in Computational Basis

Suppose we measure each qubit in the computational basis. That is, we have two random variables $X = \sigma_{ZA}$ on qubit A and $Y = \sigma_{ZB}$ on qubit B , where σ_Z is the Pauli-Z operator.

(a) Probability of measuring qubit A in state $|0\rangle$ and qubit B in $|1\rangle$:

$$p(|0\rangle, |1\rangle) \equiv p(|01\rangle) = |\langle 01| + +\rangle|^2 = \frac{1}{4}.$$

This probability can also be expressed in terms of the eigenvalues as $p(1, -1)$.

(b) Conditional probability of measuring qubit A in $|0\rangle$ given qubit B is in $|1\rangle$:

$$p(|0\rangle_A || |1\rangle_B) = p(|0\rangle_A) = \frac{1}{2}, \text{ due to independence.}$$

(c) Joint probability distribution of X and Y :

$$p(X, Y) = p(X) \cdot p(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \text{ for each outcome in } \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}.$$

(d) Conditional probability distribution of X given Y :

$$p(X|Y) = p(X) = \frac{1}{2}, \text{ for each outcome in } \{|0\rangle, |1\rangle\}.$$

(e) Probability of either qubit A being in $|0\rangle$ or qubit B being in $|1\rangle$:

$$p(|0\rangle \cup |1\rangle) = \frac{3}{4}, \text{ corresponding to the three outcomes } \{|00\rangle, |01\rangle, |11\rangle\}.$$

(f) Expected values (where $X = \sigma_{ZA}$ and $Y = \sigma_{ZB}$):

$$\langle X \rangle = p(|0\rangle)(1) + p(|1\rangle)(-1) = 0.$$

$$\langle XY \rangle = p(|00\rangle)(1 \cdot 1) + p(|01\rangle)(1 \cdot (-1)) + p(|10\rangle)((-1) \cdot 1) + p(|11\rangle)(1 \cdot 1) = 0.$$

$$\langle X + Y \rangle = p(|00\rangle)(1+1) + p(|01\rangle)(1-1) + p(|10\rangle)(-1+1) + p(|11\rangle)(-1-1) = 0.$$

Since measuring an observable X on system A and an observable Y on system B are independent events, $p(X, Y) = p(X) \cdot p(Y)$, $p(X|Y) = p(X)$, and $p(X \cup Y) = p(X) + p(Y) - p(X, Y)$.

(2) Measuring in $\{|+\rangle, |-\rangle\}$ Basis

Now we measure each qubit in the $\{|+\rangle, |-\rangle\}$ basis. The corresponding random variables are $X = \sigma_{XA}$ on qubit A and $Y = \sigma_{XB}$ on qubit B , where σ_X is the Pauli- X operator.

(a) Probability of measuring qubit A in state $|+\rangle$ and qubit B in $|-\rangle$:

$$p(|+\rangle, |-\rangle) = |\langle + - | + + \rangle|^2 = 0.$$

(b) Conditional probability of measuring qubit A in $|+\rangle$ given qubit B is in $|-\rangle$:

$$p(|+\rangle_A || |-\rangle_B) \text{ is undefined since qubit } B \text{ cannot be in } |-\rangle.$$

(c) Joint probability distribution of X and Y :

$$p(X, Y) = p(X) \cdot p(Y) = 1 \text{ for outcome } |++\rangle \text{ and } 0 \text{ for } \{|+-\rangle, |-+\rangle, |--\rangle\}.$$

(d) Conditional probability distribution of X given Y :

$p(X|Y)$ is defined only if Y is $|+\rangle$, in which case, $p(X|Y) = 1$ for X -outcome of $|+\rangle$, and 0 for $|-\rangle$.

(e) Probability of either qubit A being in $|+\rangle$ or qubit B being in $|-\rangle$:

$$p(|+\rangle \cup |-\rangle) = 1, \text{ because qubit } A \text{ is definitively in } |+\rangle.$$

(f) Expected values (where $X = \sigma_{XA}$ and $Y = \sigma_{XB}$):

$$\langle X \rangle = p(|+\rangle)(1) + p(|-\rangle)(-1) = 1.$$

$$\langle XY \rangle = p(|++\rangle)(1 \cdot 1) + \dots = 1.$$

$$\langle X + Y \rangle = p(|++\rangle)(1 + 1) + \dots = 2.$$



This example demonstrates that quantum probabilities depend on the basis (or observable) of measurement, and that conditional probabilities may be undefined for specific outcomes.

(3) Calculating Probabilities from Density Operators

In straightforward cases like this example, probabilities can be directly computed. For more complex scenarios, however, we often rely on density operators to determine these probabilities. To built intuition, let's re-calculate some of the above probabilities and expected values using density operators.

The density operator of the composite system is given by $\rho_{AB} = |++\rangle\langle++|$. The density operator of qubit A is $\rho_A = \text{tr}_B(\rho_{AB}) = |+\rangle\langle+|$. Similarly, for ρ_B . Since this is a product state, $\rho_{AB} = \rho_A \otimes \rho_B$.

$$(a) p(|0\rangle, |1\rangle) = \langle 01|\rho_{AB}|01\rangle = \frac{1}{4}.$$

$$(b) p(|+\rangle, |+\rangle) = \langle ++|\rho_{AB}|++\rangle = 1.$$

$$(c) p(|+\rangle, |-\rangle) = \langle +-|\rho_{AB}|+-\rangle = 0.$$

$$(d) p(|-\rangle_B) = \langle -|\rho_B|- \rangle = 0. \text{ Hence, } p(|+\rangle_A ||-\rangle_B) = \frac{p(|+\rangle, |-\rangle)}{p(|-\rangle_B)}$$

is undefined.

(e) Expected values:

$$\langle \sigma_{ZA} \rangle = \text{tr}(\sigma_Z \rho_A) = 0.$$

$$\langle \sigma_{ZA} \sigma_{ZB} \rangle = \text{tr}(\sigma_{ZA} \sigma_{ZB} \rho_{AB}) = 0. \quad (\sigma_{ZA} \sigma_{ZB} \equiv \sigma_{ZA} \otimes \sigma_{ZB})$$

$$\langle \sigma_{ZA} + \sigma_{ZB} \rangle = \text{tr}((\sigma_{ZA} I_B + I_A \sigma_{ZB}) \rho_{AB}) = 0.$$

$$\langle \sigma_{XA} \rangle = \text{tr}(\sigma_X \rho_A) = 1.$$

$$\langle \sigma_{XA} \sigma_{XB} \rangle = \text{tr}(\sigma_{XA} \sigma_{XB} \rho_{AB}) = 1.$$

$$\langle \sigma_{XA} + \sigma_{XB} \rangle = \text{tr}((\sigma_{XA} I_B + I_A \sigma_{XB}) \rho_{AB}) = 2.$$

Probabilities in Entangled Systems

Unlike the product state in Example 13.5, the entanglement embodied in states leads to outcomes that are intertwined in a way that is not possible in classical physics. We will use an example to illustrate how entanglement leads to inseparability of the joint state and to probabilities that exhibit non-classical correlations.

■ **Example 13.6 — Joint and Conditional Probabilities in a Bell State.** Consider two qubits A and B in the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. The joint and reduced density operators are examined in Example 13.4.

(1) Local Measurements in Computational Basis

Local measurements of each qubit in the computational basis are equivalent to observing the outcomes of the Pauli- Z operator applied to each qubit. We define two random variables $X = \sigma_{ZA}$ for qubit A and $Y = \sigma_{ZB}$ for qubit B .

(a) Probability of measuring both qubits in state $|0\rangle$:

$$p(|0\rangle, |0\rangle) = |\langle 00|\Phi^+\rangle|^2 = \frac{1}{2}.$$

(b) Joint probability distribution of X and Y :

$$p(X, Y) = [\frac{1}{2}, 0, 0, \frac{1}{2}] \text{ for outcomes } [|00\rangle, |01\rangle, |10\rangle, |11\rangle].$$



Note that $p(X, Y) \neq p(X)p(Y)$. This demonstrates that the joint state of an entangled system is inseparable into individual states, leading to probabilities that cannot be factored as in classical independent events.

- (c) Conditional probabilities given qubit B is in state $|1\rangle$:

First, $p(|1\rangle_B) = \langle 1|\rho_B|1\rangle = \frac{1}{2}$. Then,

$$p(|0\rangle_A || |1\rangle_B) = \frac{p(|01\rangle)}{p(|1\rangle_B)} = 0,$$

$$p(|1\rangle_A || |1\rangle_B) = \frac{p(|11\rangle)}{p(|1\rangle_B)} = 1.$$

Without the condition of qubit B in $|1\rangle$, we would have $p(|0\rangle_A) = p(|1\rangle_A) = \frac{1}{2}$.



These conditional probabilities demonstrate correlation. The measurement outcome of one qubit affects the probabilities of outcomes for the other qubit.

- (d) Expected values:

$$\langle X \rangle = p(|0\rangle)(1) + p(|1\rangle)(-1) = 0. \text{ Or equivalently, } \langle \sigma_{ZA} \rangle = \text{tr}(\sigma_Z \rho_A) = 0.$$

$$\langle XY \rangle = p(|00\rangle)(1 \cdot 1) + p(|01\rangle)(1 \cdot (-1)) + p(|10\rangle)((-1) \cdot 1) + p(|11\rangle)(1 \cdot 1) = 1. \text{ Or equivalently, } \langle \sigma_{ZA} \sigma_{ZB} \rangle = \text{tr}(\sigma_{ZA} \sigma_{ZB} \rho_{AB}) = 1. \text{ Note that, in this case, } \langle XY \rangle \neq \langle X \rangle \langle Y \rangle, \text{ indicating correlation.}$$

$$\langle X + Y \rangle = p(|00\rangle)(1+1) + p(|01\rangle)(1-1) + p(|10\rangle)(-1+1) + p(|11\rangle)(-1-1) = 0.$$

(2) Joint Measurements in the Bell Basis

The Bell basis, $\{|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle\}$, is an orthonormal basis for two-qubit states (see § 8.2). Measuring both qubits jointly in this basis yields a probability distribution of 1 for $|\Phi^+\rangle$ and 0 for all other basis states. ■

Exercise 13.5 Consider a two-qubit system composed of 50% in the uniform mixed state $\frac{1}{4}I$ and 50% in the Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. We perform local measurements of each qubit in the computational basis which is equivalent to observing the outcomes of the Pauli-Z operator (σ_Z) applied to each qubit. We define two random variables $X = \sigma_{ZA}$ for qubit A and $Y = \sigma_{ZB}$ for qubit B . Calculate:

- The density operator of the system ρ_{AB} .
- The reduced density operators ρ_A and ρ_B .
- The joint probability of measuring both qubits in state $|0\rangle$: $p(|0\rangle, |0\rangle)$.
- The conditional probabilities given qubit B is in state $|1\rangle$: $p(|0\rangle_A || |1\rangle_B)$ and $p(|1\rangle_A || |1\rangle_B)$.
- The expectation values $\langle X \rangle$, $\langle Y \rangle$, $\langle XY \rangle$, and $\langle X + Y \rangle$.

Key Takeaways

In the quantum realm, multiple random variables are represented by observables, whose non-commutative nature influences their measurements. Entanglement in composite systems introduces complex dependencies between variables, making quantum probabilities differ fundamentally from classical counterparts. The concept of marginal probability is extended through reduced density operators, enabling focus on specific subsystems. The nature of quantum probability is further nuanced by the distinction between product and entangled states, with the latter showing non-classical correlations.

13.1.5 Random Processes

A random or stochastic process is essentially a collection of random variables, each indexed by time or space, that describe the evolution of systems in a stochastic, or probabilistic, manner. This concept is central in understanding systems that evolve unpredictably over time or space.

1 Classical Stochastic Processes

Classical stochastic processes are mathematical models used to describe classical systems that evolve over time in a way that is at least partially random. Here are some common examples:

- Random Walk: This is a fundamental stochastic process where an object moves step by step, with each step being random. In its simplest form, at each time step, the object moves either one unit up or one unit down with certain probabilities. Random walks are used to model various phenomena, including stock market fluctuations and particle movements in liquids.
- Markov Chains: A Markov chain is a stochastic process where the probability of moving to the next state depends only on the current state and not on the sequence of events that preceded it. Markov chains are used in a variety of fields, including economics, game theory, and biology.

Random walk is a special case of Markov chain where each step is taken randomly and independently, typically with equal probability, in a specific state space like a lattice or a grid.

- Queueing Models: Used extensively in operations research, these models study the behavior of queues (or lines). They help in understanding and predicting queue lengths and waiting times, important in designing and managing facilities like call centers, hospitals, and manufacturing plants.

2 Quantum Stochastic Processes

Quantum stochastic processes extend these concepts into the quantum domain, featuring unique quantum mechanical principles like superposition and entanglement. For example:

- Quantum Walk: Analogous to classical random walk but in the quantum realm, quantum walk exhibits features such as superposition and entanglement, leading to behavior distinct from their classical counterparts.

- **Quantum Markov Processes:** Quantum Markov Processes, as the quantum counterparts of classical Markov processes, characterize the evolution of quantum states by extending the classical notion of memorylessness to quantum systems. In these processes, the progression of a quantum state is depicted through a sequence of Completely Positive Trace-Preserving (CPTP) maps. These maps ensure each transition from one state to another is independent of prior states or transitions, thereby maintaining the Markov property. Furthermore, CPTP maps can incorporate effects such as measurements, noise, and decoherence, important in accurately representing quantum systems.
- **Quantum Queueing Models:** Emerging in the context of quantum communication and computing, these models apply quantum mechanical principles to queueing theory. Quantum Queueing Models can be used for the efficient management of quantum information processes, such as the transmission of qubits in quantum networks and the execution of operations in quantum computing. They deal with challenges unique to quantum systems, like the no-cloning theorem and quantum entanglement, offering new ways to optimize quantum resource management and information processing.

3 Simulation Techniques in Quantum Stochastic Processes

Simulations help us understand quantum stochastic processes by offering tools to model and study complex quantum behaviors where analytical solutions are often intractable. Among the main methods employed are:

- **Quantum Monte Carlo (QMC) Simulations:** QMC methods use stochastic sampling to solve quantum mechanical problems, particularly valuable in studying many-body systems. These simulations help calculate properties like ground state energies and molecular structures.

In QMC, random samples are used to estimate the properties of a quantum system, effectively turning a complex quantum problem into a statistical one. Techniques like Variational Monte Carlo (VMC) and Diffusion Monte Carlo (DMC) employ different strategies for handling the probabilistic aspects of quantum mechanics.

- **Time-Dependent Schrödinger Equation:** This fundamental equation in quantum mechanics describes how the quantum state of a physical system changes over time. While QMC employs random sampling, the Schrödinger equation models the probabilistic evolution of quantum states over time.
- **Quantum Operations Sequence:** This approach models quantum stochastic processes through a sequence of quantum operations, including measurements and unitary transformations.

13.1.6 *Example: One-Dimensional Quantum Walk

In this subsection, we delve into the one-dimensional quantum walk as a representative example of quantum probability and quantum stochastic processes. We will contrast its characteristics with those of classical random walks, highlighting the distinctive features of quantum mechanics that come into play in the quantum variant.

1 Classical 1D Random Walk

The one-dimensional random walk is a simple yet instructive model. It can be visualized as a walk along a line, with each step being of the same length. Starting at the origin, a fair (unbiased) coin is flipped before each step. A heads (H) results in a step forward, while tails (T) leads to a step backward.

The problem of interest is determining the probability of being at a specific point n after N steps, given that the starting point is 0. We denote this probability as $f_N(n)$.

Since the walker must be at some position after N steps, the sum of these probabilities for all possible positions n must equal 1, i.e., $\sum_n f_N(n) = 1$.

The pattern in these probabilities becomes clearer when factoring out $\frac{1}{2^N}$:

n	-5	-4	-3	-2	-1	0	1	2	3	4	5
$f_0(n)$						1					
$2f_1(n)$					1		1				
$2^2 f_2(n)$				1		2		1			
$2^3 f_3(n)$			1		3		3		1		
$2^4 f_4(n)$		1		4		6		4		1	
$2^5 f_5(n)$	1		5		10		10		5		1

To better understand these probabilities, one can enumerate the sequences of coin flips leading to each position. For instance, in a three-step walk ($N = 3$), the sequence HHH will end at $n = 3$, while HHT, HTH, and THH will end at $n = 1$. The negative positions correspond to the reverse combinations of H and T. With $2^3 = 8$ possible three-step walks, the probabilities for the different positions are: $f_3(-3) = 1/8$, $f_3(-1) = 3/8$, $f_3(1) = 3/8$, $f_3(3) = 1/8$.

These probabilities correspond to the coefficients in Pascal's Triangle, and they are the same as those in the binomial expansion of $(a+b)^N$. Therefore, the probability function is given by:

$$f_N(n) = \frac{1}{2^N} \binom{N}{\frac{N+n}{2}} = \frac{N!}{2^N (\frac{N+n}{2})! (\frac{N-n}{2})!}, \quad (13.29)$$

where n is an integer such that $-N \leq n \leq N$ and $N + n$ is even.

2 Quantum Simulation of Classical 1D Random Walk

We can simulate a classical 1D random walk using a quantum circuit, as illustrated in Fig. 13.2.

The first qubit represents the state of a fair coin, denoted as $|\text{coin}\rangle$. The second register of d qubits records the position of the walker, denoted as $|\text{pos}\rangle$. This register is initialized to $|100\dots0\rangle$, representing position $n = 0$. The offset ensures that $n = 0$ is centered in the set $\{0, 1\}^d$. We choose a sufficiently large d such that $2^d > 2N$.

The state of the $d + 1$ qubits is given by

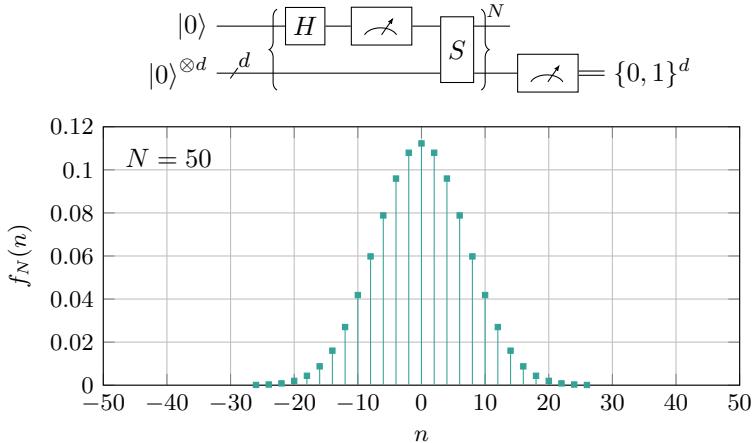


Figure 13.2: Quantum Simulation of 1D Random Walk

$$|\Psi\rangle = |\text{coin}\rangle \otimes |\text{pos}\rangle. \quad (13.30)$$

A quantum random number generator (QRNG, see § 5.4) simulates the coin flip. A single-bit QRNG is implemented as a qubit initialized in the state $|0\rangle$ or $|1\rangle$, rotated by a Hadamard gate, followed by a measurement in the computational basis. After each measurement, the qubit state collapses to $|0\rangle$ or $|1\rangle$, so the QRND is ready for use again.

We construct a quantum gate S , acting on the $d+1$ qubits, such that if $|\text{coin}\rangle = |0\rangle$ (representing Heads), n is incremented; if $|\text{coin}\rangle = |1\rangle$ (representing Tails), n is decremented:

$$S = |0\rangle\langle 0| \otimes \sum_n |n+1\rangle\langle n| + |1\rangle\langle 1| \otimes \sum_n |n-1\rangle\langle n|. \quad (13.31)$$

Exercise 13.6 Verify that the shift operator S , given in Eq. 13.31, is unitary.

We repeat the QRNG and S operations N times and measure the last d qubits. The probability of finding the d -qubit state corresponding to position n yields the function $f_N(n)$:

$$f_N(n) = |\langle n|\text{pos}\rangle|^2. \quad (13.32)$$

The results of a simulation with $N = 50$ and $d = 7$ are graphed in the lower part of Fig. 13.2. The graph mirrors Eq. 13.29, a binomial probability distribution centered around $n = 0$ with a spread of \sqrt{N} .

3 1D Quantum Walk

Finally, we are ready to explore the 1D quantum walk. Unlike a classical random walk, a quantum walk is fundamentally deterministic in its evolution, with probabilistic outcomes emerging from quantum measurement. For this reason, we omit “random” in its name.

Consider a quantum particle as the walker, initially at the origin on a number line. In each step of a quantum walk, the particle enters a superposition of moving one step left and one step right. After multiple steps, the particle does not merely occupy a single position; rather, it has a probability amplitude for every position it could have reached. These amplitudes from different steps interfere, leading to a probability distribution distinct from classical random walks. Notably, some off-center positions may exhibit higher probabilities than the center position $n = 0$ due to constructive interference.

The quantum circuit for simulating this phenomenon is shown in Fig. 13.3. The key difference from the classical walk simulation (Fig. 13.2) is that we no longer measure the QRNG after each step. This lack of measurement allows the system state to maintain superposition and demonstrate interference effects.

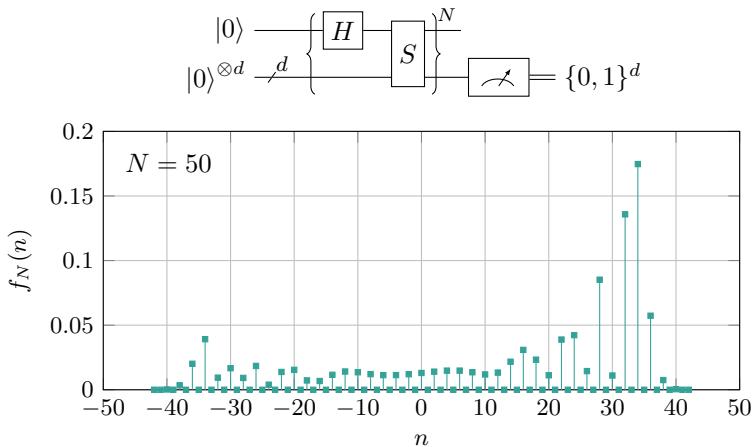


Figure 13.3: Quantum 1D Walk

A simulation with $N = 50$ and $d = 7$ produces a result graphed in the lower part of Fig. 13.3. Unlike the bell-shaped binomial distribution of the classical walk, the quantum walk yields a distinctly shaped distribution, exhibiting higher probabilities for positions further from the start point due to quantum interference.

To elucidate this behavior, we demonstrate the evolution of the system state under Eq. 13.31 in Table 13.1, started with the coin state $|0\rangle$. Noticeably, by the third step, interference shifts the probability distribution, making it asymmetrical. This asymmetry arises because the terms $-|1, -1\rangle + |1, -1\rangle$ cancel each other, eliminating their probability contribution at $x = -1$. Conversely, the terms $|0, 1\rangle + |0, 1\rangle$ combine to form $2|0, 1\rangle$, contributing a probability of $\frac{4}{8}$.

It is important to note that the initial coin state significantly influences the shape of the $f_N(n)$ curve. This initial state can be a general state represented as $\alpha|0\rangle + \beta|1\rangle$.

Exercise 13.7 Construct a table showing the evolution of system states for a 1D quantum walk over the first three steps, following the dynamics described by Eq. 13.31. Assume the initial state of the coin is $|1\rangle$. Your table should be similar to Table 13.1, detailing the coin and walker states, as well as the probability

Steps	Coin & Walker State	Probability
0	$ 0, 0\rangle$	[1]
1	$\frac{1}{\sqrt{2}}(0, 1\rangle + 1, -1\rangle)$	$[\frac{1}{2}, 0, \frac{1}{2}]$
2	$\frac{1}{2}(0, 2\rangle + 1, 0\rangle + 0, 0\rangle - 1, -2\rangle)$	$[\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4}]$
3	$\begin{aligned} &\frac{1}{\sqrt{8}}(0, 3\rangle + 1, 1\rangle + 0, 1\rangle - 1, -1\rangle \\ &+ 0, 1\rangle + 1, -1\rangle - 0, -1\rangle + 1, -3\rangle) \\ &= \frac{1}{\sqrt{8}}(0, 3\rangle + 1, 1\rangle + 2 0, 1\rangle - 0, -1\rangle + 1, -3\rangle) \end{aligned}$	$[\frac{1}{8}, 0, \frac{1}{8}, 0, \frac{5}{8}, 0, \frac{1}{8}]$

Table 13.1: Evolution of Coin & Walker States in a 1D Quantum Walk

distribution at each step.

4 Applications

Although the one-dimensional quantum walk has been introduced primarily as an illustrative example of quantum probability and quantum stochastic processes, the generalized concept of quantum walks is a rich and complex topic with significant applications in quantum computing and quantum information theory. These applications exploit the unique properties of quantum walks, including superposition and interference. The following are some key areas where quantum walks are applied:

- **Quantum Algorithms:** Quantum walks are a powerful tool for designing quantum algorithms. They can explore computational spaces more efficiently than their classical counterparts due to quantum superposition and interference. Algorithms based on quantum walks have been shown to offer speedups for certain problems. For example, the quantum walk-based search algorithm can search an unsorted database quadratically faster than any possible classical algorithm.
- **Quantum Simulation:** Quantum walks can simulate various quantum systems, particularly in studying transport phenomena in quantum systems. They are instrumental in simulating the behavior of electrons in materials, which helps in understanding and designing new quantum materials and devices.
- **Quantum Computing Models:** The concept of quantum walks underpins certain models of quantum computation, like the Continuous-Time Quantum Walk (CTQW) model. These models provide alternative frameworks to the standard gate-based quantum computing model and are more natural for certain types of problems and physical implementations.
- **Quantum Communications:** In the realm of quantum communication, quantum walks are used to model and analyze quantum networks, potentially leading to more efficient protocols for quantum information transfer. Quantum walks also have applications in developing new cryptographic protocols.

13.2 Quantum Entropy and Information

The concept of entropy, a cornerstone in information theory, measures the uncertainty

or information content associated with a random variable. In the quantum regime, this notion extends to encapsulate the uncertainty in quantum states.



In information theory, “information” is defined in a way that is quite different from its everyday usage. It’s not about the message, meaning, or knowledge, but rather about the uncertainty or unpredictability of a message source. For instance, if a message is very predictable, it contains less information than a highly unpredictable one. This is quantified using the concept of entropy, usually represented in bits. The higher the entropy, the more information a message carries.

13.2.1 Classical Shannon Entropy

Consider a classical random variable X with a sample space Ω . Each outcome x in Ω occurs according to a probability distribution $p(X)$. The information content (or self-information, to distinguish with joint information) $I(x)$ of a particular outcome x is defined as:

$$I(x) = -\log p(x). \quad (13.33)$$

The logarithmic base determines the unit of information; for base 2, the unit is bits.

The log function in the definition ensures that information is additive. Consider two independent events x and y with probabilities $p(x)$ and $p(y)$, respectively. The joint probability of both events occurring is the product $p(x)p(y)$. This relationship leads to the total information content for both events being $I(x, y) = -\log(p(x)p(y)) = I(x) + I(y)$.

The Shannon entropy is defined as the expected value (statistical average) of the information content of a random variable over all possible outcomes:

$$H(X) = E[I(x)] = - \sum_{x \in \Omega} p(x) \log p(x). \quad (13.34)$$

Thus, in information theory, “information” is quantified as the reduction in uncertainty about a random variable and is measured by the change in entropy, when a message is received.

■ **Example 13.7 — Roll of a Six-sided Fair Die.** Consider the roll of a six-sided fair die, with $\Omega = \{1, 2, 3, 4, 5, 6\}$ and a uniform probability distribution for each outcome. The Shannon entropy in this case is $H(X) = - \sum_{x=1}^6 \frac{1}{6} \log \frac{1}{6} = \log 6 \approx 2.58$ bits. This value represents the average uncertainty or information content associated with the outcome of a roll. In information theory, it indicates that, on average, about 2.58 bits is needed to optimally encode the outcome of a die roll in a large series of rolls. ■



In the context of information theory and probability simulation, we often use the phrase “draw from probability distribution p ”. For example, consider a simple probability distribution p over $\Omega = \{a, b\}$, where $p(a) = 0.6$, $p(b) = 0.4$. A “draw from p ” means randomly selecting either a or b , with a 60% chance of selecting a .

and a 40% chance for b in any single draw.

■ **Example 13.8 — Simulating a Draw from a Distribution.** Suppose we have a probability distribution p over a finite set of outcomes. In a computational simulation, $H(p)$ represents the least number of truly random bits (coin flips) needed on average to generate a draw from p . This is especially accurate when simulating a large number of draws, as the average number of bits used converges to $H(p)$.

For instance, consider $p = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ for the sample space $\{a, b, c\}$. To generate a draw from p using minimum number of coin flips, if we get Heads (with a probability of $\frac{1}{2}$), we select a , done; if we get Tails, we flip a second coin, and select b or c according to Heads or Tails (each with overall probability of $\frac{1}{4}$). Thus, on average, it takes $1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.5$ coin flips to generate a draw. This is consistent with the Shannon entropy calculation, $H(p) = 1.5$ bits. ■

■ **Example 13.9 — Storing a Draw with Optimal Compression.** For a probability distribution p , the average number of bits required to store a draw from p , using the best possible compression scheme, is also given by $H(p)$. This stems from the fact that Shannon entropy defines the limit of lossless compression. For each draw from p , one can store the sequence of random bits used to generate that particular outcome. Over many such draws, the average length of these stored bit sequences will approximate $H(p)$.

As an example, consider again $p = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ for the sample space $\{a, b, c\}$. To encode a string drawn from this distribution, we use the shortest code 0 for the most frequent letter a , and then 10 for b , 11 for c . On average, we need 1.5 bits per letter. This is consistent with the Shannon entropy calculation, $H(p) = 1.5$ bits. ■

Exercise 13.8 Calculate the entropy in bits for each of the following random variables, representing pixel values in an image:

- Assume we have an image where each pixel's color is represented by a unique combination of Red, Green, and Blue channels, each of which can take any integer value from 0 to 255 with uniform probability. Calculate the entropy of a single pixel.
- Now we remove the red color in the image, i.e., the Red channel is always zero, and only the Green and Blue channels vary from 0 to 255 with uniform probability. Calculate the entropy of a single pixel under this condition.
- Finally, assume the color space of the image is limited: the Red channel varies from 0 to 63, the Green channel from 0 to 127, and the Blue channel from 64 to 255, all with uniform probability. Calculate the entropy of a single pixel for this scenario.

Note: Assume each color channel is independent of the others.

Basic Properties

Shannon entropy has a number of important properties:

- $0 \leq H(p) \leq \log(d)$, where d is the dimension of the sample space.

2. $H(p) = 0$ if $p = 1$ for some event and 0 for the rest, representing a deterministic process.
3. $H(p) = \log(d)$ if p is uniform, $(\frac{1}{d}, \dots, \frac{1}{d})$, representing maximum randomness.

13.2.2 *Quantum von Neumann Entropy

In quantum probability theory, observables are a non-commutative generalization of random variables, with their set of eigenvalues playing the role of the set of possible outcomes. A given density operator ρ generalizes the role of a probability distribution by completely encapsulating the probabilistic state of a quantum system, allowing us to compute statistical quantities such as the expected value of an observable A :

$$\langle A \rangle = \text{tr}(\rho A) = \langle \rho, A \rangle. \quad (13.35)$$

To compute the von Neumann entropy, which serves as a measure of uncertainty in quantum systems, analogous to Shannon entropy in classical probability, we consider $A = -\log \rho$ as an observable. The von Neumann entropy $S(\rho)$ of a quantum system described by a density operator ρ is defined as:

$$S(\rho) = -\langle \log \rho \rangle = -\text{tr}(\rho \log \rho). \quad (13.36)$$

This expression can be simplified using the spectral decomposition of ρ . In the eigenbasis of ρ , where ρ is diagonal, the entropy takes the form:

$$S(\rho) = -\sum_i \lambda_i \log \lambda_i, \quad (13.37)$$

where each eigenvalue λ_i represents the probability of the quantum system being in the corresponding eigenstate. This result is invariant under basis changes, thanks to the basis-independence of the trace operation.

The von Neumann entropy quantifies the amount of information present in a quantum system. It has a number of properties similar to those of Shannon entropy:

1. $0 \leq S(\rho) \leq \log(d)$, where d is the dimension of ρ .
2. $S(\rho) = 0$ if some term in $\{\lambda_i\}$ is 1 and the rest 0, representing a pure state.
3. $S(\rho) = \log(d)$ if $\{\lambda_i\}$ is uniform, representing a maximumly mixed state.

■ **Example 13.10 — Pure States.** For the pure state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the density operator is:

$$\rho = \begin{bmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{bmatrix}.$$

The eigenvalues of ρ are 0 and 1. Therefore, the von Neumann entropy for ρ is:

$$S(\rho) = 0 \log 0 + 1 \log 1 = 0. \quad (13.38)$$

Here we take $0 \log 0 = 0$ in the limiting sense: $\lim_{x \rightarrow 0^+} x \log x = 0$.

In fact, the entropy for any pure state of any dimension, including Bell states, is 0, reflecting the absence of uncertainty or information entropy in a pure state. This

is because $\rho = |\psi\rangle\langle\psi|$ of a pure state $|\psi\rangle$ is a rank-1 matrix, i.e., with only 1 linearly independent column vector or row vector. For such a matrix, one of the eigenvalues is always 1 (due to the normalization of quantum states), and all others are 0. ■

■ **Example 13.11 — The Uniform Mixed State.** For a d -dimensional qudit, the density operator for the uniform mixed state is:

$$\rho = \frac{1}{d} I_{d \times d}.$$

This density operator is already diagonal. Its von Neumann entropy is:

$$S(\rho) = - \sum_{i=1}^d \frac{1}{d} \log \frac{1}{d} = \log d. \quad (13.39)$$

Exercise 13.9 A machine randomly produces qubits in a mixed state, where the state is $|0\rangle$ with probability p and $|1\rangle$ with probability $1-p$ for $0 \leq p \leq 1$. Define the density matrix ρ of this mixed state. Compute the von Neumann entropy $S(\rho)$ of the qubits. Determine the value of p that maximizes $S(\rho)$, and find the maximum value of $S(\rho)$.

13.2.3 *Classical Joint Entropy and Mutual Information

Building on our understanding of Shannon entropy, this subsection delves into information measures involving multiple random variables. We introduce and examine the concept of mutual information, which plays a key role in understanding the interdependencies and shared information between random variables. We will focus on the case of two random variables, with the principles being extendable to more complex scenarios involving additional variables. These measures not only broaden our perspective on information theory but also lay the groundwork for their quantum counterparts.

1 Joint Entropy

In classical information theory, given two random variables X and Y , the joint entropy, $H(X, Y)$, is defined in terms of the Shannon entropy of the joint probability distribution $p(x, y)$ (see § 13.1.4):

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y), \quad (13.40)$$

which quantifies the total uncertainty in the joint system of X and Y , encompassing all combinations of random events x and y .

Since $p(x, y) = p(y, x)$, it follows that $H(X, Y)$ is symmetric, i.e., $H(X, Y) = H(Y, X)$.

If X and Y are independent, then $p(x, y) = p(x)p(y)$, resulting in:

$$H(X, Y) = H(X) + H(Y). \quad (13.41)$$

2 Conditional Entropy

The conditional entropy of Y given X is defined in terms of the conditional probability distribution $p(y|x)$. For a specific outcome $X = x$, the conditional entropy is determined using the Shannon formula:

$$H(Y|x) = - \sum_{y \in Y} p(y|x) \log p(y|x). \quad (13.42)$$

The full conditional entropy of Y given X is defined as the expected value of $H(Y|x)$ over all possible outcomes of X :

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|x) \quad (13.43a)$$

$$= - \sum_{x \in X} \sum_{y \in Y} p(x)p(y|x) \log p(y|x) \quad (13.43b)$$

$$= - \sum_{x \in X} \sum_{y \in Y} p(y,x) \log p(y|x), \quad (13.43c)$$

which represents the average uncertainty in Y given knowledge of X , averaged over all possible outcomes of X .

Similarly, the conditional entropy of X given Y is:

$$H(X|Y) = - \sum_{y \in Y} \sum_{x \in X} p(x,y) \log p(x|y). \quad (13.44)$$

3 Relationships

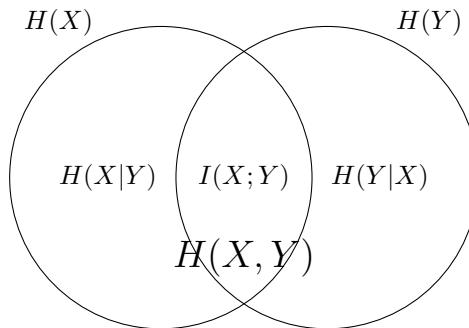


Figure 13.4: Venn Diagram for Shannon Entropy Relationships

Due to the chain rule in probability (Eq. 13.19):

$$p(x,y) = p(y,x) = p(x|y)p(y) = p(y|x)p(x), \quad (13.45)$$

joint and conditional entropies satisfy the identity:

$$H(X,Y) = H(Y,X) = H(Y|X) + H(X) = H(X|Y) + H(Y). \quad (13.46)$$

Or, equivalently:

$$H(Y|X) = H(Y, X) - H(X), \quad (13.47a)$$

$$H(X|Y) = H(X, Y) - H(Y). \quad (13.47b)$$

These identities illustrate that the conditional entropy measures the additional uncertainty about one variable when the other is known. These relationships can be illustrated using a Venn diagram as in Fig. 13.4.



In Fig. 13.4, the exclusive region of $H(X)$ relative to $H(Y)$ is represented by $H(X) - I(X; Y)$, equal to $H(X|Y)$.

Conversely, in Fig. 13.1, the exclusive part of $p(x)$ relative to $p(y)$ is shown as $p(x) - p(x, y)$, which differs from $p(x|y)$.

Readers are encouraged to explore the reasons behind this distinction.

4 Mutual Information

Mutual information is defined as a measure of the amount of information that one random variable contains about another. Mathematically, it is given by:

$$I(X; Y) = H(X) + H(Y) - H(X, Y). \quad (13.48)$$

In this notation, the semicolon (instead of a comma) is used to emphasize that mutual information is a measure of the dependence between the two variables, rather than a joint distribution of them.

The mutual information $I(X; Y)$ quantifies the shared information between two random variables, X and Y . It represents the reduction in uncertainty about one variable given knowledge of the other. Specifically, if X and Y are independent, $I(X; Y) = 0$, indicating that knowledge of X provides no information about Y , and vice versa. It has been established that $I(X; Y) \geq 0$ for all cases, which implies that gaining knowledge about one variable cannot decrease our information about another; such knowledge can only maintain or increase the amount of shared information.

In the Venn diagram in Fig. 13.4, $I(X; Y)$ is the intersection between the two circles representing $H(X)$ and $H(Y)$, while $H(X, Y)$ is their union. Notably, both $I(X; Y)$ and $H(X, Y)$ are symmetric: $I(X; Y) = I(Y; X)$ and $H(X, Y) = H(Y, X)$.

Exercise 13.10 Verify the following equivalent expressions for mutual information:

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (13.49a)$$

$$= H(X) - H(X|Y) \quad (13.49b)$$

$$= H(Y) - H(Y|X) \quad (13.49c)$$

$$= \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \quad (13.49d)$$

Mutual information also exhibits additivity for independent variables under certain conditions. For instance, if the joint distribution of two pairs of random variables (X, Y) and (W, Z) can be factorized as $p(X, Y, W, Z) = p(X, Y)p(W, Z)$, indicating that (X, Y) is independent of (W, Z) , then the mutual information between these two sets of variables is additive: $I(X, W; Y, Z) = I(X; Y) + I(W; Z)$.

■ **Example 13.12 — Perfectly Correlated Systems.** Suppose that X is a random variable whose entropy $H(X)$ is 8 bits, and that $Y(X)$ is a deterministic function that takes on a different value for each value of X .

- The entropy of Y , $H(Y)$, is also 8 bits because Y is a one-to-one function of X .
- The conditional entropy of Y given X , $H(Y|X) = 0$, since X uniquely determines Y .
- The conditional entropy of X given Y , $H(X|Y) = 0$, because Y uniquely determines X due to the function being one-to-one.
- The joint entropy $H(X, Y)$ is $H(X) + H(Y|X) = 8$ bits, which is just the entropy of X since Y provides no additional uncertainty.
- The mutual information $I(X; Y) = H(X) + H(Y) - H(X, Y) = 8$, consistent with the fact that knowing X provides full information about Y and vice versa.
- Thus, for two perfectly correlated systems,

$$I(X; Y) = H(X, Y) = H(X) = H(Y), \text{ and}$$

$$H(X|Y) = H(Y|X) = 0.$$

■

■ **Example 13.13 — Partially Correlated Systems.** Similar to Example 13.12, X is a random variable whose entropy $H(X)$ is 8 bits. Now, $Y(X)$ is a deterministic function but is no longer invertible; that is, different values of X may correspond to the same value of $Y(X)$.

- Since now different values of X may correspond to the same value of $Y(X)$, the new distribution of Y has lost entropy because it has less variability and therefore $H(Y) < 8$ bits.
- The conditional entropy of Y given X , is still $H(Y|X) = 0$ bits, since X still uniquely determines Y .
- Now, knowledge of Y no longer uniquely determines X , and so the conditional entropy $H(X|Y)$ is no longer zero because there is some uncertainty about X even after knowing Y : $H(X|Y) > 0$.
- The joint entropy $H(X, Y)$ is still $H(X) + H(Y|X) = 8$ bits, which is just the entropy of X since Y provides no additional uncertainty. Even though $H(X|Y) > 0$ and $H(Y) < 8$, we still have $H(Y) + H(X|Y) = 8$.
- The mutual information $I(X; Y) = H(X) + H(Y) - H(X, Y) = H(Y) < H(X)$, measuring the amount of information shared between X and Y which is now smaller than 8 bits.
- Thus, for two partially correlated systems,

$$H(X), H(Y) \leq H(X, Y) \leq H(X) + H(Y),$$

$$0 \leq H(X|Y) \leq H(X), 0 \leq H(Y|X) \leq H(Y), \text{ and}$$

$$0 \leq I(X; Y) \leq H(X), H(Y).$$

■

■ **Example 13.14 — Independent Systems.** Similar to Example 13.12, X is a random variable whose entropy $H(X)$ is 8 bits. But now $Y(X)$ is completely

random, i.e., given a value of $X = x$, Y takes on a random value y from a probability distribution with the same entropy as X .

- (a) We are given $H(Y) = H(X) = 8$ bits, even though Y and X are now independent.
- (b) The conditional entropy of Y given X , is $H(Y|X) = 8$ bits, since X does not influence Y .
- (c) Similarly, the conditional entropy of X given Y , is $H(X|Y) = 8$ bits as well.
- (d) The joint entropy $H(X, Y)$ is $H(X) + H(Y|X) = H(Y) + H(X|Y) = 16$ bits.
- (e) The mutual information $I(X; Y) = H(X) + H(Y) - H(X, Y) = 0$, signifying that no information is shared between X and Y .
- (f) Thus, for two independent systems,

$$H(X, Y) = H(X) + H(Y),$$

$$H(X|Y) = H(X), H(Y|X) = H(Y), \text{ and}$$

$$I(X; Y) = 0.$$

Exercise 13.11 Draw the corresponding Venn diagram for each of the Examples 13.12 to 13.14.

13.2.4 *Quantum Joint Entropy and Mutual Information

Having established a solid foundation in both the von Neumann entropy and classical information measures, we now turn our attention to quantum information measures involving two random variables. This generalization provides a deeper insight into how quantum mechanics redefines informational relationships and interdependencies.

Given a bipartite quantum system AB described by a density operator ρ_{AB} which encodes the probabilities of the system, the pair of random variables X and Y are replaced by ρ_{AB} . The two subsystems are described by the reduced density operators $\rho_A = \text{tr}_B(\rho_{AB})$ and $\rho_B = \text{tr}_A(\rho_{AB})$. (See § 12.2.8 for details.)

1 Joint and Conditional Entropies

The standard definitions of joint and conditional von Neumann entropies adopt the form of the classical relations, with $S(\rho_{AB})$ in place of the classical joint entropy, $S(\rho_A)$ substituted for $H(X)$, and $S(\rho_B)$ substituted for $H(Y)$. Thus,

$$S(A, B) \equiv S(\rho_{AB}) = -\text{tr}(\rho_{AB} \log \rho_{AB}), \quad (13.50a)$$

$$S(A) \equiv S(\rho_A) = -\text{tr}(\rho_A \log \rho_A), \quad (13.50b)$$

$$S(B) \equiv S(\rho_B) = -\text{tr}(\rho_B \log \rho_B). \quad (13.50c)$$

Since conditional density operators are rarely used, the conditional von Neumann entropies are defined directly according to the classical entropy relationship Eq. 13.47:

$$S(A|B) = S(\rho_{AB}) - S(\rho_B), \quad (13.51a)$$

$$S(B|A) = S(\rho_{AB}) - S(\rho_A). \quad (13.51b)$$

2 Mutual Information

Mutual information is defined similarly to the classical case, but with quantum entropies. For two quantum systems A and B , the mutual information $I(A; B)$ is given by:

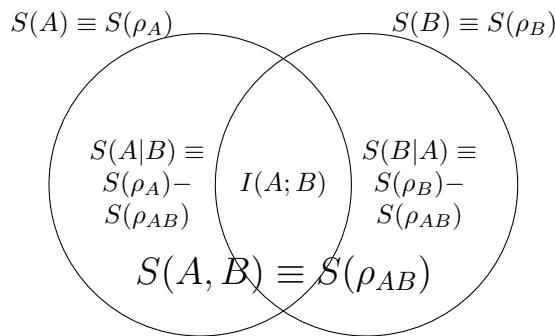
$$I(A; B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}). \quad (13.52)$$

The terms $S(\rho_A)$ and $S(\rho_B)$ represent the individual uncertainties of systems A and B , respectively, while $S(\rho_{AB})$ represents the uncertainty of the joint system. The mutual information $I(A; B)$ quantifies how much the knowledge of one system (say A) reduces the uncertainty about the other system (B), thus signifying the total amount of correlation between two quantum systems, reflecting the interconnectedness of their quantum states.

Notably, since ρ_A and ρ_B are derived from ρ_{AB} through partial trace, the joint and conditional entropies, as well as the mutual information, are all encoded in ρ_{AB} .

3 Relationships

The relationships among quantum entropies and information can be depicted using a Venn diagram, as shown in Fig. 13.5. However, it's important to note that such diagrams cannot capture the phenomenon of negative quantum conditional entropy, which occurs in entangled states.



Note: Quantum conditional entropy can be negative in the case of entangled states, which is not captured by the diagram. Refer to the text for details.

Figure 13.5: Venn Diagram for von Neumann Entropy Relationships

In the case where the systems A and B are independent, $\rho_{AB} = \rho_A \otimes \rho_B$, $S(\rho_{AB}) = S(\rho_A) + S(\rho_B)$, and $I(A; B) = 0$.

4 Negative Conditional Entropy

A surprising aspect of quantum conditional entropy is that it can be negative for entangled states, a phenomenon that has no counterpart in classical information theory. This is exemplified in Example 13.15.

In classical systems, conditional entropy quantifies the reduction in uncertainty about one part of the system when another part is known, and it is always non-negative. This reflects that knowledge about a system can only decrease or leave unchanged the uncertainty associated with its parts. Conversely, quantum systems

exhibit a unique property where conditional entropy can become negative. This occurs due to quantum entanglement, indicating that the interconnections between subsystems are so profound that knowing the state of the whole system actually implies a greater certainty (less uncertainty) than the sum of what is known about the individual parts. Such negative conditional entropy is exclusive to quantum mechanics, underscoring the fundamentally non-classical behavior introduced by entanglement.

However, even though quantum conditional entropy can be negative, quantum mutual information is still always non-negative: $I(A; B) \geq 0$. (A proof of this inequality is beyond the scope of this text.)

■ **Example 13.15 — Conditional Entropy of an EPR Pair.** Consider the archetypal case where Alice and Bob share an EPR pair, $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Its joint density operator is $\rho_{AB} = |\Phi^+\rangle\langle\Phi^+|$.

As derived in Example 13.4, the reduced density operator for both Alice and Bob is $\rho_A = \rho_B = \text{tr}_B(\rho_{AB}) = \text{tr}_A(\rho_{AB}) = I/2$, representing maximally mixed states.

Consequently, the entropies are $S(\rho_{AB}) = 0$ for the pure joint state, and $S(\rho_A) = S(\rho_B) = 1$ for the maximally mixed reduced states. This leads to a surprising result in the conditional entropies:

$$\begin{aligned} S(A|B) &= S(\rho_{AB}) - S(\rho_B) = -1, \\ S(B|A) &= S(\rho_{AB}) - S(\rho_A) = -1. \end{aligned}$$

The negative values of $S(A|B)$ and $S(B|A)$ signify the presence of entanglement and the departure from classical probabilistic correlations.

The mutual information is

$$I(A; B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) = 2,$$

signifying that the total amount of entanglement-based correlation is equivalent to 2 bits of information.

This is in contrast to classical systems, where the mutual information between two perfectly correlated random variables X and Y would be $I(X; Y) = H(X) = H(Y)$. The factor of 2 in $I(A; B) = 2S(\rho_A)$ for Bell states indicates that quantum correlations from entanglement can be stronger than any classical correlation.

In general, if ρ_{AB} represents a pure state with $S(\rho_{AB}) = 0$, then it holds that $S(\rho_A) = S(\rho_B)$. For a product state, where ρ_{AB} is separable into ρ_A and ρ_B , both are pure states and hence $S(\rho_A) = S(\rho_B) = 0$. Conversely, for a maximally entangled state like a Bell state, $S(\rho_A) = S(\rho_B)$ attains its maximum value of 1. Thus, $S(\rho_A) = S(\rho_B)$ is often referred to as the *entanglement entropy* for a pure state, quantifying the degree of entanglement. ■

Exercise 13.12 Consider a two-qubit mixed state which is composed of an EPR pair $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ with 50% probability, and another entangled state $|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ with 50% probability.

Compute the joint and conditional von Neumann entropies for this mixed state.

Key Takeaways

In quantum information theory, von Neumann entropy fulfills the roles of the classical Shannon entropy. Joint entropy in bipartite quantum systems, derived from the joint density operator ρ_{AB} , quantifies the total uncertainty associated with the combined state of the two systems. Mutual information, derived from the joint and reduced density operators of these systems, measures the amount of information shared between them, reflecting the degree of their quantum correlations. Notably, quantum conditional entropy can be negative in entangled systems, which underscores the non-classical nature of quantum entanglement, a distinctive feature absent in classical information theory.

13.3 * Core Theorems in Quantum Information

In the preceding chapters, we have explored two theorems fundamental to the field of quantum computation and information theory. The no-cloning theorem, as outlined in § 5.1.3, asserts that it is impossible to create an identical copy of an arbitrary unknown quantum state. This theorem highlights a fundamental restriction on quantum information imposed by quantum mechanics. In addition, the no-communication theorem (§ 9.7) establishes that superluminal communication, or faster-than-light transmission of information, is not feasible, despite the presence of non-local correlations inherent in quantum entanglement.

This section ventures into more advanced territories of quantum communications. We will explore two significant theorems that further shape our understanding of quantum information—the quantum data processing inequality and the Holevo theorem—as well as the property of strong subadditivity of von Neumann entropy. Our focus will be on elucidating the background and implications of these theorems and principles, offering insights into their roles in the broader landscape of quantum information science. While the mathematical underpinnings of these concepts are profound, we aim to emphasize their conceptual significance and practical implications, thereby making these advanced topics accessible to a broader audience.

Prerequisite Review: Quantum Channels

This topic is detailed in § 12.2.7. Below is a brief summary of the relevant concepts.

A Completely Positive Trace-Preserving (CPTP) map is an operation that transforms one density operator into another while preserving trace and positivity. It includes both unitary and non-unitary transformations, is applicable to closed and open quantum systems, and allows for changes in dimensionality.

CPTP maps are often modeled using quantum channels, also known as Kraus decompositions. A quantum channel can be expressed as:

$$\mathcal{E}(\rho) = \sum_j K_j \rho K_j^\dagger, \quad (13.53)$$

where K_j are the Kraus operators that describe the quantum channel. These operators embody the transformations the system undergoes, and the sum of their effects ensures the preservation of the trace and positivity of the density operator. The specific dynamics of the channel determine the form and number of the Kraus operators.

13.3.1 *The Data Processing Inequality

The Data Processing Inequality (DPI) is a fundamental concept in information theory, both classical and quantum. It dictates how information is transformed and preserved within systems undergoing various processes. This subsection explores the DPI, focusing first on its quantum formulation.

1 The Quantum Data Processing Inequality

The von Neumann entropy is invariant under any unitary transformation U :

$$S(U\rho U^\dagger) = S(\rho). \quad (13.54)$$

This invariance signifies that entropy remains constant through reversible information processing in closed quantum systems.

For more general transformations, quantum systems undergo completely positive, trace-preserving (CPTP) maps, also known as quantum channels in quantum information theory. A CPTP map \mathcal{E} transforms a density operator ρ into another density operator $\mathcal{E}(\rho)$, preserving the positivity and total probability. (See § 12.2.7 for more details.)

The von Neumann entropy does not decrease under CPTP maps [66]:

$$S(\mathcal{E}(\rho)) \geq S(\rho). \quad (13.55)$$

This relationship (which encompasses Eq. 13.54) is known as the quantum Data Processing Inequality (DPI). The proof of this inequality is mathematically sophisticated, involving advanced concepts in quantum theory, and falls outside the purview of this discussion. For a detailed exploration, readers are encouraged to consult the references provided at the end of this chapter.

■ **Example 13.16 — Measuring a Bell State.** Suppose we have a Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ with zero entropy. Upon measuring one qubit in the computational basis, the state collapses to a mixed state, either $|00\rangle$ or $|11\rangle$ with equal probability, resulting in an entropy of 1. ■

■ **Example 13.17 — Bit-Flip Noise.** Consider a qubit state $|\psi\rangle = |0\rangle$, which initially has zero entropy:

$$\rho = |0\rangle\langle 0|, \quad S(\rho) = 0. \quad (13.56)$$

A bit-flip noise channel \mathcal{E} acts on each qubit with probability p , flipping $|0\rangle$ to $|1\rangle$ and vice versa:

$$\mathcal{E}(\rho) = (1 - p)\rho + p\sigma_X\rho\sigma_X, \quad (13.57)$$

where σ_X is the Pauli- X (bit-flip) operator. The resulting state $\mathcal{E}(\rho)$ is a mixed state with increased entropy:

$$\mathcal{E}(\rho) = (1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|, \quad (13.58a)$$

$$S(\mathcal{E}(\rho)) = -[(1-p)\log(1-p) + p\log(p)]. \quad (13.58b)$$

$S(\mathcal{E}(\rho))$ has a maximum value of 1 at $p = 0.5$. ■

Key Takeaways

The quantum DPI indicates that the entropy of a quantum system cannot decrease under the application of CPTP operations. These operations encompass quantum noise, decoherence, and other non-unitary processes commonly encountered in open quantum systems, as well as unitary operations in closed quantum systems.

2 Implications

The quantum DPI is fundamental in understanding the behavior of quantum systems, especially in the context of quantum information theory and quantum thermodynamics.

The increase in von Neumann entropy under CPTP maps reflects several fundamental aspects of quantum systems. This includes their tendency to lose coherence and approach an equilibrium characterized by maximally mixed states, indicative of maximal entropy and reduced quantum information. Additionally, it highlights the irreversibility of certain quantum processes, the inherent uncertainties in quantum states, and the dynamics of quantum entanglement, particularly in multipartite systems where increased local entropy can signal entanglement. From an information-theoretic perspective, this entropy increase also represents a degradation in the system capacity to retain and process quantum information.

Quantum thermodynamics is an interdisciplinary field that merges the principles of quantum mechanics with those of thermodynamics. It extends traditional thermodynamics to the quantum realm, exploring how quantum properties like superposition and entanglement influence and are influenced by thermodynamic processes. Key areas of interest include understanding the thermodynamics of small systems at quantum scales, energy exchanges in quantum systems, the thermodynamic cost of quantum information processes, and the second law of thermodynamics in quantum systems. This field provides important insights for developing quantum technologies like quantum computers, quantum batteries, and quantum heat engines.

In this context, quantum DPI is often compared to the second law of thermodynamics because both describe a non-decreasing property of the system entropy. In classical thermodynamics, the second law states that the total entropy of a closed system does not decrease over time, reflecting the natural tendency towards disorder or equilibrium in isolated systems. However, while the second law is a statement about closed classical systems, the quantum DPI often applies to open quantum systems undergoing CPTP operations.

3 Classical DPI Versus Quantum DPI

In classical information theory, the behavior of entropy under different types of transformations diverges from quantum systems (refer to Examples 13.12 to 13.14):

1. When using deterministic and invertible functions g , the entropy $H(X)$ remains unchanged, i.e., $H(g(X)) = H(X)$. In this scenario, each outcome of X uniquely corresponds to an outcome of $g(X)$ and vice versa, preserving the level of uncertainty in the system.
2. With deterministic but non-invertible functions g , entropy may decrease ($H(g(X)) \leq H(X)$). This reduction occurs because non-invertible functions can map multiple outcomes of X into a single outcome of $g(X)$, thereby diminishing the uncertainty and, consequently, the entropy.
3. In the case of stochastic functions g , we observe an increase in entropy ($H(g(X)) \geq H(X)$) due to the additional randomness introduced by g , which heightens the system's uncertainty.

Thus, the classical counterpart of the quantum DPI, namely $H(g(X)) \geq H(X)$, is not universally applicable. Deterministic but non-invertible functions in classical systems can actually decrease the entropy through many-to-one mappings. In contrast, general quantum transformations are described by CPTP maps, which inherently ensure that the von Neumann entropy does not decrease.

13.3.2 *The Holevo Bound and Channel Capacity

This subsection explores a foundational theorem (the HSW theorem) that defines the capabilities of quantum channels in information transmission. It focuses on Holevo's theorem and its implications for channel capacity, providing key insights into the potential and limitations of quantum communication networks.

1 Background

Imagine Alice wishes to send information, encoded in a classical random variable X , to Bob. Using n classical bits, she can transmit a maximum of n bits of information. But what happens if she employs qubits instead? Unlike a classical bit, which represents two discrete states, a pure qubit state can represent any point on the Bloch sphere. For instance, Alice could encode a random n -bit string $s \in \{0, 1\}^n$ into a single qubit state as:

$$|\psi\rangle_x = \cos \frac{\pi x}{N} |0\rangle + \sin \frac{\pi x}{N} |1\rangle, \quad (13.59)$$

where x is the decimal representation of the bit string and $N = 2^n$.

When Alice sends a qubit to Bob, either directly or through quantum teleportation, it is natural to question whether Bob can fully recover x (or the n -bit string s). Unfortunately, the answer is negative. To derive classical information from the qubit, Bob must perform a measurement. Upon measuring the qubit, Bob will find it in state $|0\rangle$ with probability $p_0 = \cos^2(\frac{\pi x}{N})$ and in state $|1\rangle$ with probability $1 - p_0$. To accurately determine $\frac{\pi x}{N}$, and thus the information encoded by Alice, Bob would need to measure a large number of identically prepared qubits. However, the quantum no-cloning theorem prohibits creating exact copies of an unknown quantum state, which precludes Bob from generating the necessary multiple qubits for precise measurement. Alice could send Bob multiple copies, but the very act would significantly reduce the information transmission efficiency. Consequently, Alice's capacity to encode a wealth of information into a single qubit state does not translate into Bob's ability to retrieve it, primarily due to the probabilistic nature of quantum measurements and the constraints of quantum cloning.

The Holevo Bound addresses this issue by quantifying the maximum efficiency of information transmission using qubits. Surprisingly, the Holevo Bound dictates that an n -qubit stream can convey at most n bits of classical information. This bound confirms that irrespective of the quantum information encoded (as represented by the states of the qubits), an observer is limited to extracting at most n bits of classical information from the measurement of these n qubits.

The optimal efficiency of classical information transmission using a qubit is achieved only when the information is encoded using a set of orthogonal states. Given that a qubit has only two orthogonal states, it can transmit at most one bit of classical information per measurement. However, when encoding information with non-orthogonal states, or in scenarios involving quantum correlations, the amount of accessible classical information per qubit may be less than 1 bit.

Thus, the Holevo Bound demonstrates that despite a qubit's ability to exist in a continuum of states due to superposition, it cannot convey more than 1 bit of classical information to an observer performing any permissible quantum measurement.

In superdense coding (see § 10.2), it is shown that a qubit can transmit two bits of information, achievable through the pre-sharing of an entangled pair. Nevertheless, this still equates to two bits of classical information being transmitted using two qubits.

This result, known as the Holevo-Schumacher-Westmoreland (HSW) theorem, sets an upper limit on the amount of classical information transmissible via a quantum system.

2 The HSW Theorem

The Holevo-Schumacher-Westmoreland (HSW) theorem is anchored in the scenario of a quantum communication channel, wherein a sender (Alice) aims to transmit classical information to a receiver (Bob) via quantum states. In this setup, X denotes the classical information Alice seeks to convey, encoded within quantum states $\{\rho_x\}$, while Y encapsulates the classical information Bob decodes from his measurements of these states. The mutual information, $I(X; Y)$, quantifies the reduction in uncertainty about X (Alice's message) afforded by knowledge of Y (Bob's measurement outcomes), representing the *accessible information*.

Preceding the HSW theorem, the Holevo Bound addresses a related yet simpler query: given a collection of quantum states $\{\rho_x\}$, each with a probability $p(x)$, what is the maximal quantity of classical information that can be obtained from measurements? The Holevo Bound posits that the accessible information, $I(X; Y)$, is constrained by:

$$I(X; Y) \leq S \left(\sum_x p(x) \rho_x \right) - \sum_x p(x) S(\rho_x), \quad (13.60)$$

where $S(\rho)$ signifies the von Neumann entropy of the state ρ . This limitation is met when the quantum states $\{\rho_x\}$ are mutually orthogonal, such as $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$.

The HSW theorem extends the Holevo Bound to establish the limits of information transmission through quantum channels. As depicted in Fig. 13.6, the HSW theorem builds on this by considering:

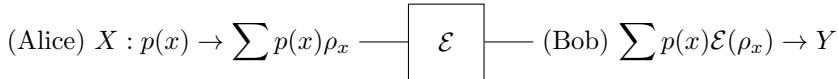


Figure 13.6: Schematic of the Quantum Information Transmission Process

- Alice encoding classical information from a random variable X with a probability distribution $p(x)$ into a set of quantum states $\{\rho_x\}$.
- These quantum states are transmitted to Bob via a quantum channel described by a Completely Positive Trace-Preserving (CPTP) map \mathcal{E} .
- Bob performs measurements to decode the information, which results in a new random variable Y containing the classical information he extracts.

The HSW theorem posits that the mutual information $I(X; Y)$ is bounded by the Holevo channel capacity χ , formulated as:

$$\chi = \max_{\{p(x), \rho_x\}} \left(S\left(\sum_x p(x)\mathcal{E}(\rho_x)\right) - \sum_x p(x)S(\mathcal{E}(\rho_x)) \right), \quad (13.61)$$

where the maximization runs over all ensembles of states $\{\rho_x\}$ and their probability distribution $p(x)$.

This theorem sets an upper bound on the classical capacity of a quantum channel, illustrating that quantum mechanics, despite allowing for elaborate state representations, limits the classical information extractable from a quantum system. It also emphasizes the role of quantum entropy in determining the capacity of quantum channels.

The HSW theorem is a cornerstone in the field of quantum information theory. Its implications are significant, particularly in quantum cryptography and the creation of efficient quantum communication protocols.

Implications of Holevo's Bound on Quantum Computing

A system of n qubits can encode quantum information in a composite quantum state represented by 2^n complex numbers. This expansive encoding capacity underlies the remarkable parallel computational power of quantum computing. However, transitioning from quantum computations to retrievable classical results necessitates measurements, introducing inherent constraints. Holevo's bound limits the amount of information that can be extracted to n classical bits.

Quantum algorithms, such as Deutsch-Jozsa or Shor's, ingeniously exploit quantum interference to efficiently condense the outcome of quantum computations into these n bits. This condensation process ensures that critical information from the quantum computation is compacted into a form suitable for output measurements. These algorithms showcase the delicate balance between maximizing quantum processing capabilities and navigating the fundamental constraints of information extraction during the quantum-to-classical transition.

The derivation of the HSW theorem is complex, rooted in the depths of quantum information theory, and extends beyond the introductory nature of this text. Those

seeking a comprehensive understanding are advised to refer to the literature listed at the end of this chapter.

3 Revisiting the Holevo Bound

Having explored the theoretical underpinnings of the HSW theorem, let's revisit our initial discussion about the Holevo Bound. To understand why a quantum channel cannot transmit more than 1 bit of classical information, we analyze the Holevo channel capacity, as expressed in Eq. 13.61. The first term represents the entropy of the average state of the ensemble after passing the channel. For a single qubit, this is at most 1 bit when the state is maximally mixed. The second term is the weighted average of the entropies of the individual quantum states. When each state $\mathcal{E}(\rho_x)$ is pure, its entropy is 0, and consequently, the second term sums to 0.

Therefore, the maximal value of χ is achieved when the average state of the ensemble is maximally mixed, yielding an entropy of 1 bit, and simultaneously, each individual state is pure, contributing no entropy. This scenario results in $\chi_{\max} = 1$ bit.

13.3.3 * Subadditivity and Strong Subadditivity

The subadditivity and strong subadditivity properties of von Neumann entropy mathematically encapsulate how entropy behaves in quantum operations, forming a crucial underpinning for various key theorems. Notably, these properties are instrumental in the proofs of the quantum data processing inequality (DPI) and the Holevo-Schumacher-Westmoreland (HSW) theorem, both of which are central to our understanding of quantum communication and information processing.

1 Subadditivity (SA)

For any composite quantum system AB , subadditivity asserts that the total entropy of the system does not exceed the sum of the entropies of its constituent subsystems. Mathematically, this is expressed as:

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B), \quad (13.62)$$

where ρ_{AB} is the density operator of the composite system AB , and ρ_A and ρ_B are the reduced density operators of the subsystems A and B , respectively.

This principle of subadditivity indicates that the total entropy of combined quantum systems is constrained by the sum of the individual entropies of each system.

A related inequality that sets the lower bound of $S(\rho_{AB})$ is the Araki-Lieb inequality:

$$S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)| \quad (13.63)$$

An important implication of subadditivity is that in a quantum system, the mutual information—a measure of the total correlations between subsystems A and B —is always non-negative:

$$I(A; B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \geq 0. \quad (13.64)$$

2 Strong Subadditivity (SSA)

Strong Subadditivity (SSA) of von Neumann entropy further generalizes the concept of subadditivity to tripartite quantum systems, encompassing three subsystems A , B , and C . The SSA is mathematically expressed as:

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}), \quad (13.65)$$

where ρ_{ABC} is the density operator for the combined system and ρ_{AB}, ρ_{BC} are the density operators for two of the bipartite subsystems.

This inequality encapsulates a more intricate interplay of entropic relationships in a composite system. It implies that the total entropy of the entire system ABC plus the entropy of one of its parts (B) is no greater than the sum of the entropies of two overlapping subsystems (AB and BC).

While the proof of the classical version of SSA is relatively straightforward, proving the quantum SSA presents significant challenges. This increased difficulty arises from the non-commutative nature of the density operators that describe quantum subsystems. Despite these challenges, several distinct proofs have been developed to establish SSA in quantum systems.

SSA is a powerful tool in quantum information, providing a deeper understanding of entropy dynamics in complex quantum systems and forming a theoretical basis for key theorems in quantum communication and information processing.

13.4 Further Exploration

For further exploration of topics in quantum information, the following resources are recommended:

- Michael A Nielsen and Isaac L Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. ISBN: 978-1-107-00217-3.
- John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. ISBN: 978-1-107-18056-7.
- Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2017. ISBN: 978-1-107-17616-4.

These references provide comprehensive insights into quantum computation, information theory, and the intricate dynamics of quantum stochastic processes.

13.5 Summary and Conclusions

Chapter Overview

This chapter on quantum information delves into three core areas: quantum probability, quantum entropy and information, and advanced theorems. It starts by contrasting quantum probability with classical probability, highlighting unique quantum features like entanglement. The discussion then shifts to quantum entropy, comparing it with classical Shannon entropy, and examines quantum mutual information. Finally, the chapter explores advanced concepts like the quantum

data processing inequality, the Holevo theorem, and subadditivity in quantum systems, emphasizing their importance in quantum communication and theoretical underpinnings.

Quantum Probability Essentials

The first section delves into the essentials of quantum probability, highlighting its distinctive features compared to classical probability. Key concepts such as complex-valued probability amplitudes, entanglement, and non-commutativity of quantum observables are explored. The section provides a foundational understanding of how quantum probability resembles and differs from classical probability. This exploration lays the groundwork for understanding quantum entropy and information.

Quantum Entropy and Information

Next, the chapter shifts focus to quantum entropy and information. It provides a comprehensive understanding of classical Shannon entropy and its quantum counterpart, von Neumann entropy, highlighting their roles in quantifying uncertainty in classical and quantum systems. Additionally, the concepts of classical joint entropy, mutual information, and their quantum analogs are explored. Through these discussions, the section offers insights into how quantum mechanics reshapes the understanding of informational relationships and interdependencies, demonstrating the unique nature of quantum information.

Core Theorems in Quantum Information

The final part of the chapter delves into core theorems and concepts central to quantum information theory. It begins with an exploration of the quantum data processing inequality, which posits that the entropy of a quantum system cannot decrease under the application of CPTP operations. This principle underlines the non-decreasing nature of entropy in quantum processes, reflecting phenomena such as the loss of coherence, the irreversibility of certain quantum operations, and the constraints on information extraction from quantum systems.

The discussion then moves to the Holevo theorem, which sets fundamental limits on the amount of classical information that can be transmitted using quantum states through quantum channels. This theorem is essential in understanding the boundaries of quantum communication.

Additionally, this section covers the properties of subadditivity and strong subadditivity of von Neumann entropy. These mathematical formulations are useful in analyzing the entropic characteristics of composite quantum systems.

This part of the chapter weaves together these advanced concepts, providing insights into their theoretical importance and practical implications in the broader landscape of quantum information science.

Problem Set 13

13.1 (Review of classical probability and information.) Consider random variables,

X and Y , each taking values in $\Omega = \{1, 2, \dots, 6\}$, representing the outcomes of rolling fair dice. We impose the condition that the sum $X + Y$ is odd. Your tasks are:

- (a) Find the probability distributions of X , Y , (X, Y) , and $X + Y$.
- (b) Compute the joint entropy $H(X, Y)$.
- (c) Compute the individual entropies $H(X)$ and $H(Y)$.
- (d) Compute the conditional entropies $H(Y|X)$ and $H(X|Y)$.
- (e) Describe what each of these quantities represents in the context of this problem.
- (f) Verify Eq. 13.46 using your computations.

13.2 For classical Shannon entropy, we have the property that $H(X) \leq H(X, Y)$, reflecting the intuition that the uncertainty of a single variable X cannot exceed that of the joint system X and Y . This exercise explores whether a similar property holds in the quantum realm with von Neumann entropy, specifically, whether $S(\rho_A) \leq S(\rho_{AB})$ for the following cases:

- (a) A Bell state $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.
- (b) A classically correlated state $\rho_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)$.

13.3 Coherent information is defined as $I(A|B) = S(\rho_B) - S(\rho_{AB})$. It is known to be a measure of quantum correlations. Positivity of coherent information indicates that quantum correlations are present. Use the following cases to demonstrate this.

- (a) A Bell state $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.
- (b) A product state $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle)$.

13.4 Consider a mixed state described by the density operator ρ . Since ρ is Hermitian, it can also be regarded as an observable. Set $H = \rho$ in Eq. 13.14 for expected value and Eq. 13.15 for the variance. Simplify these equations and discuss their physical meanings, particularly in the context of the purity of the state.

13.5 Consider a two-qubit (A and B) system composed of 50% in the uniform mixed state $\frac{1}{4}I$ and 50% in the Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. We define two random variables $U = \sigma_X \otimes I$ and $V = I \otimes \sigma_Z$, where σ_X and σ_Z are Pauli operators. They represent measuring qubit A in the $\{|+\rangle, |-\rangle\}$ basis, and qubit B in the computational basis.

- (a) Calculate the density operator of the two-qubit system ρ_{AB} .
- (b) Verify that U and V commute.
- (c) Find all the common eigenstates for U and V , e.g., $|+0\rangle \equiv |+\rangle \otimes |0\rangle$.
- (d) Calculate the probability distribution on set of common eigenstates for U and V .
- (e) Calculate the conditional probabilities given qubit B is in state $|1\rangle$: $p(|+\rangle_A || 1\rangle_B)$ and $p(|-\rangle_A || 1\rangle_B)$.

- (f) Calculate the expectation values $\langle U \rangle$, $\langle V \rangle$, $\langle UV \rangle$, and $\langle U + V \rangle$.
- (g) Calculate the von Neumann entropy of the two-qubit system.
- (h) Calculate the reduced density operators ρ_A and ρ_B .
- (i) Calculate the joint entropy $S(\rho_{AB})$, and the entropies of the subsystems $S(\rho_A)$ and $S(\rho_B)$.
- (j) Calculate the mutual information $I(A; B)$.

13.6 Consider two quantum systems A and B , each in a maximally mixed state of dimension d :

$$\rho_A = \rho_B = \frac{1}{d} \sum_{i=1}^d |i\rangle\langle i|.$$

The joint system is described by the density operator:

$$\rho_{AB} = \frac{1}{d} \sum_{i=1}^d |i\rangle\langle i| \otimes |i\rangle\langle i|.$$

Prove that the entropies $S(\rho_A)$ and $S(\rho_B)$ are both $\log d$, and that the joint entropy $S(\rho_{AB})$ and mutual information $I(A; B)$ are also $\log d$, indicating perfect classical correlation in the joint system.

13.7 Consider a mixed state represented by the density operator $\rho = p|0\rangle\langle 0| + (1 - p)|1\rangle\langle 1|$, where the state is $|0\rangle$ with probability p and $|1\rangle$ with probability $1 - p$ for $0 \leq p \leq 1$.

Apply the following CPTP map representing a depolarizing channel:

$$\mathcal{E}(\rho) = (1 - s)\rho + \frac{s}{2}I,$$

where s is the probability of depolarization, and I is the identity operator.

Your tasks are:

- (a) Calculate the von Neumann entropy of the initial state ρ .
- (b) Apply the depolarizing channel \mathcal{E} to ρ and find the resulting state.
- (c) Calculate the von Neumann entropy of the state after applying the map.
- (d) Compare the entropies before and after applying \mathcal{E} to illustrate that entropy does not decrease.

13.8 * Prove the Data Processing Inequality (DPI) from the Strong Subadditivity (SSA) property.

The data processing inequality in quantum information theory asserts that the entropy S of a quantum state ρ cannot decrease under the action of a Completely Positive Trace-Preserving (CPTP) map \mathcal{E} . Formally, this is expressed as:

$$S(\mathcal{E}(\rho)) \geq S(\rho).$$

To prove the data processing inequality using the strong subadditivity of quantum entropy,

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}),$$

follow these steps:

- a. Consider a tripartite system described by the state ρ_{ABC} , with ρ_A being the subsystem of interest and ρ_{BC} an environment subsystem used in the Stinespring dilation for a CPTP map \mathcal{E} (see § 12.2.7.1). Here, $\rho_{ABC} = |\psi\rangle\langle\psi|_{ABC}$ represents a unitary dilation on an expanded system, and $\rho_A = \text{tr}_{BC}(\rho_{ABC})$ is the initial state of subsystem A .
- b. Apply the CPTP map \mathcal{E} to the subsystem ρ_C of ρ_{ABC} , resulting in a new state $\rho'_{ABC} = (I_A \otimes I_B \otimes \mathcal{E}_C)(\rho_{ABC})$.
- c. Trace out the environment subsystem (BC) in ρ'_{ABC} to obtain $\mathcal{E}(\rho_A) = \rho'_A$, where ρ'_A is the state of subsystem A after the application of the CPTP map, illustrating the Stinespring dilation theorem in action.
- d. Utilize the strong subadditivity property and the non-increasing nature of entropy under partial trace to connect $S(\rho_{ABC})$ and $S(\rho_{BC})$ with $S(\rho_A)$ and $S(\mathcal{E}(\rho_A))$.
- e. Apply strong subadditivity to demonstrate that $S(\mathcal{E}(\rho_A)) \geq S(\rho_A)$, thereby establishing the data processing inequality.

13.9 * Prove the Holevo Bound using Strong Subadditivity.

The Holevo Bound is a fundamental result in quantum information theory that establishes an upper limit on the amount of classical information that can be transmitted securely using quantum states. Specifically, it limits the accessible information that can be extracted from a quantum ensemble.

Given an ensemble of quantum states $\{\rho_x\}$ with probabilities $\{p(x)\}$, the Holevo Bound states that the mutual information between the classical variable X representing the choice of quantum state and the quantum variable Y representing the measurement outcomes is bounded by:

$$\chi(\{p(x), \rho_x\}) = S\left(\sum_x p(x)\rho_x\right) - \sum_x p(x)S(\rho_x),$$

where χ is the Holevo quantity, $S(\cdot)$ denotes the von Neumann entropy, and $\sum_x p(x)\rho_x$ is the average state of the ensemble.

To prove the Holevo Bound using the Strong Subadditivity (SSA) property, follow these steps:

- a. Begin by considering a quantum ensemble $\mathcal{X} = \{p(x), \rho_x\}$, where each quantum state ρ_x is prepared with probability $p(x)$. Construct a tripartite system ρ_{ABC} , where A represents the classical variable X encoded in a quantum state, B is the quantum system prepared in the state ρ_x corresponding to X , and C is an ancillary system introduced to purify the state of AB .

- b. Express the average state $\rho = \sum_x p(x) \rho_x$ and consider the purification ρ_{ABC} such that $\rho_{AB} = \sum_x p(x) |x\rangle_A \langle x| \otimes \rho_x$ and $\rho = \text{tr}_{A,C}(\rho_{ABC})$.
- c. Apply the strong subadditivity property, $S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})$, to the constructed quantum ensemble and purification, taking into account that the purification ρ_{ABC} makes $S(\rho_{ABC}) = 0$ and $S(\rho_{AB})$ represents the entropy of the classical-quantum state.
- d. Use the definition of the Holevo quantity $\chi(\mathcal{X})$ and relate it to the entropies in the strong subadditivity inequality. Specifically, show how $\chi(\mathcal{X})$ can be expressed in terms of $S(\rho_{AB})$ and $S(\rho_B)$, considering that $S(\rho_B) = S(\rho)$ due to the purification process.
- e. Conclude that the Holevo quantity $\chi(\mathcal{X})$, representing the accessible information from the quantum ensemble, is bounded by the difference $S(\rho) - \sum_x p(x)S(\rho_x)$, thereby proving the Holevo Bound.



Supporting Materials

Essential Mathematics: Quick References	442
Bibliography	455
List of Figures	464
List of Tables	465
Index	466
Journey Forward	472

Essential Mathematics: Quick References

A Complex Numbers

Basic Relations

Imaginary unit: $i \equiv \sqrt{-1}$. $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, $i^5 = i$, ...

	Cartesian Form	Exponential Form
	$z = x + iy$	$z = re^{i\theta}$
Conjugate	$z^* = x - iy$	$z^* = re^{-i\theta}$
Modulus	$ z = \sqrt{zz^*} = \sqrt{x^2 + y^2}$	$ z = r$
Conversion	$x = r \cos \theta$	$r = \sqrt{x^2 + y^2}$
	$y = r \sin \theta$	$\theta = \arctan2(y, x)$

Basic Operations

Given

$$z_1 = x_1 + iy_1 = r_1 e^{i\theta_1}, z_2 = x_2 + iy_2 = r_2 e^{i\theta_2}:$$

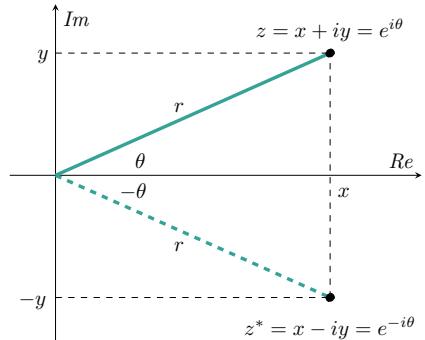
$$(z_1 \cdot z_2)^* = z_1^* \cdot z_2^* \quad |z_1 \cdot z_2| = |z_1| \cdot |z_2|$$

$$z_1 \cdot z_2 = r_1 e^{i\theta_1} \cdot r_2 e^{i\theta_2} = r_1 r_2 e^{i(\theta_1 + \theta_2)}$$

$$\frac{z_1}{z_2} = \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}$$

$$z_1 \cdot z_2 = x_1 x_2 - y_1 y_2 + i(x_1 y_2 + x_2 y_1)$$

$$\frac{z_1}{z_2} = \frac{z_1 z_2^*}{z_2 z_2^*} = \frac{x_1 x_2 + y_1 y_2 - i(x_1 y_2 - x_2 y_1)}{x_2^2 + y_2^2}$$



Useful Formulas

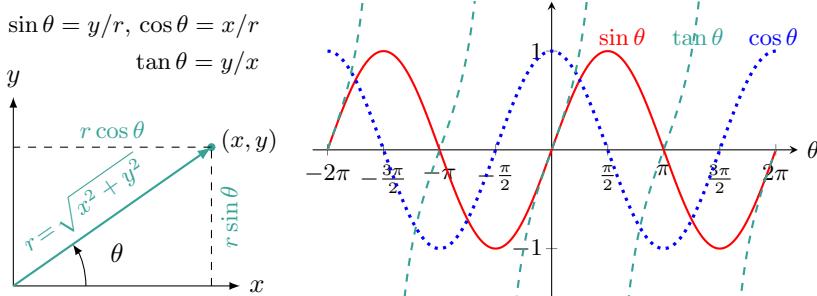
Euler's formula: $e^{i\theta} = \cos \theta + i \sin \theta$

De Moivre's formula: $(\cos \theta + i \sin \theta)^n = \cos(n\theta) + i \sin(n\theta)$

Roots of unity ($\omega^n = 1$): $\omega_k = e^{ik\frac{2\pi}{n}}$, where $k = 0, 1, \dots, n-1$. $\sum_{k=0}^{n-1} \omega_k = 0$.

B Trigonometry

Definitions and Basic Properties



	$\sin \theta$	$\cos \theta$	$\tan \theta$	$\csc \theta$	$\sec \theta$	$\cot \theta$
Definition	y/r	x/r	y/x	r/y	r/x	x/y
Period	2π	2π	π	2π	2π	π
Range	$[-1, 1]$	$[-1, 1]$	$(-\infty, \infty)$	$(-\infty, -1] \cup [1, \infty)$	$(-\infty, \infty)$	
Zeros	$n\pi$	$(n + \frac{1}{2})\pi$	$n\pi$			$(n + \frac{1}{2})\pi$
Poles			$(n + \frac{1}{2})\pi$	$n\pi$	$(n + \frac{1}{2})\pi$	$n\pi$
Inv Range	$[-\frac{\pi}{2}, \frac{\pi}{2}]$	$[0, \pi]$	$(-\frac{\pi}{2}, \frac{\pi}{2})$	$[-\frac{\pi}{2}, \frac{\pi}{2}]$	$[0, \pi]$	$(-\frac{\pi}{2}, \frac{\pi}{2})$

Note: n is an integer. $\sin \theta$ is also denoted as $\sin(\theta)$. The inverse of $x = \sin \theta$, $\theta = \arcsin x$, is also written as $\theta = \sin^{-1} x$. The extended inverse $\arcsin 2(y, x)$, or $\text{asin} 2(y, x)$, extends the range to $[-\pi, \pi]$ by considering the signs of x and y . Similarly, for other trig functions.

Special Values

θ	0	$\frac{\pi}{6}, 30^\circ$	$\frac{\pi}{4}, 45^\circ$	$\frac{\pi}{3}, 60^\circ$	$\frac{\pi}{2}, 90^\circ$	$\frac{3\pi}{4}, 135^\circ$	$\pi, 180^\circ$	$\frac{5\pi}{4}, 225^\circ$
$\sin \theta$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{2}}{2}$	0	$-\frac{\sqrt{2}}{2}$
$\cos \theta$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	$-\frac{\sqrt{2}}{2}$	-1	$-\frac{\sqrt{2}}{2}$
$\tan \theta$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	∞	-1	0	1

Interrelations

$$\tan \theta = \frac{\sin \theta}{\cos \theta} \quad \cot \theta = \frac{1}{\tan \theta} \quad \sec \theta = \frac{1}{\cos \theta} \quad \csc \theta = \frac{1}{\sin \theta}$$

Cofunction Formulas

$$\sin(\frac{\pi}{2} - \theta) = \cos \theta \quad \cos(\frac{\pi}{2} - \theta) = \sin \theta \quad \cot(\frac{\pi}{2} - \theta) = \tan \theta$$

Pythagorean Identities

$$\sin^2 \theta + \cos^2 \theta = 1 \quad \tan^2 \theta + 1 = \sec^2 \theta \quad 1 + \cot^2 \theta = \csc^2 \theta$$

Symmetry Properties

$$\begin{array}{lll} \sin(-\theta) = -\sin\theta & \sin(\pi - \theta) = \sin\theta & \sin(\pi + \theta) = -\sin\theta \\ \cos(-\theta) = \cos\theta & \cos(\pi - \theta) = -\cos\theta & \cos(\pi + \theta) = -\cos\theta \\ \tan(-\theta) = -\tan\theta & \tan(\pi - \theta) = -\tan\theta & \tan(\pi + \theta) = \tan\theta \end{array}$$

Double Angle Formulas

$$\begin{aligned} \sin 2\theta &= 2\sin\theta\cos\theta \\ \cos 2\theta &= \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta \\ \tan 2\theta &= \frac{2\tan\theta}{1 - \tan^2\theta} \end{aligned}$$

Half Angle Formulas

$$\sin^2 \frac{\theta}{2} = \frac{1 - \cos\theta}{2} \quad \cos^2 \frac{\theta}{2} = \frac{1 + \cos\theta}{2} \quad \tan \frac{\theta}{2} = \frac{\sin\theta}{1 + \cos\theta} = \frac{1 - \cos\theta}{\sin\theta}$$

Sum and Difference Formulas

$$\begin{aligned} \sin(\alpha \pm \beta) &= \sin\alpha\cos\beta \pm \cos\alpha\sin\beta & \cos(\alpha \pm \beta) &= \cos\alpha\cos\beta \mp \sin\alpha\sin\beta \\ \tan(\alpha \pm \beta) &= \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta} \end{aligned}$$

Product-to-Sum Formulas

$$\begin{array}{ll} 2\sin\alpha\sin\beta = \cos(\alpha - \beta) - \cos(\alpha + \beta) & 2\cos\alpha\cos\beta = \cos(\alpha - \beta) + \cos(\alpha + \beta) \\ 2\sin\alpha\cos\beta = \sin(\alpha + \beta) + \sin(\alpha - \beta) & 2\cos\alpha\sin\beta = \sin(\alpha + \beta) - \sin(\alpha - \beta) \end{array}$$

Sum-to-Product Formulas

$$\begin{array}{ll} \sin\alpha + \sin\beta = 2\sin\frac{\alpha + \beta}{2}\cos\frac{\alpha - \beta}{2} & \sin\alpha - \sin\beta = 2\cos\frac{\alpha + \beta}{2}\sin\frac{\alpha - \beta}{2} \\ \cos\alpha + \cos\beta = 2\cos\frac{\alpha + \beta}{2}\cos\frac{\alpha - \beta}{2} & \cos\alpha - \cos\beta = -2\sin\frac{\alpha + \beta}{2}\sin\frac{\alpha - \beta}{2} \end{array}$$

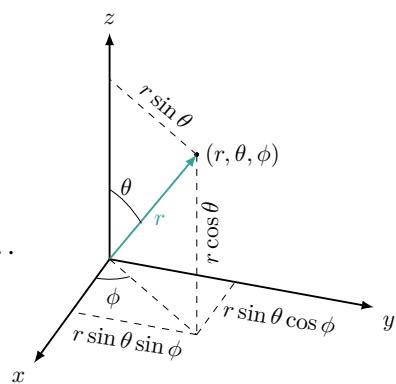
Laws of Sines and Cosines

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c} \quad a^2 = b^2 + c^2 - 2bc\cos A$$

Spherical Coordinate System

Useful Taylor Expansions

$$\begin{aligned} e^x &= \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \\ \sin x &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots \\ \cos x &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots \end{aligned}$$



C Linear Algebra for QCI

Arrays in 0D, 1D, 2D, ...

0D: scalar, c , in \mathbb{C}

1D: vector, $|a\rangle = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, in \mathbb{C}^n

2D: matrix, $A = \begin{bmatrix} a_{11} & a_{12} & \dots \\ a_{21} & a_{22} & \ddots \\ \vdots & & \end{bmatrix}$, in $\mathbb{C}^{n \times n}$

3D, 4D, ...: tensor, in $\mathbb{C}^{n \times n \times \dots \times n}$ (k times), or \mathbb{C}^{n^k}

Matrix Definitions

A^* : complex conjugate of A

A^T : transpose of A

A^\dagger : Hermitian conjugate or adjoint of A , $A^\dagger = (A^T)^*$

Hermitian (self-adjoint) matrix: $H^\dagger = H$

Unitary matrix: $U^\dagger = U^{-1}$ or $U^\dagger U = I$

Normal matrix: $AA^\dagger = A^\dagger A$ (includes Hermitian matrix and unitary matrix)

Trace: $\text{tr } A \equiv \sum_i a_{ii}$

Determinant: $\det A = |A|$

Commutator: $[A, B] \equiv AB - BA$

Anti-commutator: $\{A, B\} \equiv AB + BA$

Direct Sum: $A \oplus B$ (block diagonal matrix of A and B)

Ket, Bra, and Braket

Ket: $|v\rangle \equiv \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$

Bra: $\langle v| \equiv |v\rangle^\dagger = [v_1^* \quad v_2^* \quad v_3^*]$

Braket (inner product):

$$\langle u|v\rangle \equiv \langle u|\cdot|v\rangle = [u_1^* \quad u_2^* \quad u_3^*] \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = u_1^*v_1 + u_2^*v_2 + u_3^*v_3$$

$$\langle u|v\rangle = \langle v|u\rangle^*$$

Norm: $\|v\| = \sqrt{\langle v|v\rangle} = \sqrt{v_1^*v_1 + v_2^*v_2 + v_3^*v_3}$

Normalization: $|v\rangle \rightarrow |\hat{v}\rangle = \frac{|v\rangle}{\|v\|}$ so that $\|\hat{v}\| = 1$

Vector Matrix Product

$$A|v\rangle = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$(A|v\rangle)^\dagger \equiv \langle Av| = \langle v| A^\dagger$$

$$\langle u|A|v\rangle = [u_1^* \quad u_2^* \quad u_3^*] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$\langle u|A|v\rangle = \langle u|Av\rangle = \langle uA|v\rangle = \langle A^\dagger u|v\rangle$$

Eigenvalues and Eigenvectors

$$H|\phi_i\rangle = \lambda_i|\phi_i\rangle$$

A normal matrix allows for spectral decomposition. Its eigenvalues are complex, whereas those of Hermitian matrices are real, and anti-Hermitian matrices imaginary. The eigenvectors of these matrices form complete, orthonormal bases.

Orthonormal property: $\langle \phi_i|\phi_j\rangle = \delta_{ij}$

Completeness: $\sum_i |\phi_i\rangle\langle\phi_i| = I$

Spectral decomposition of H : $H = \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|$

Statistics

Observables in QM (A and B) are Hermitian matrices.

Expected value: $\langle A \rangle_\psi = \langle \psi|A|\psi\rangle$

Standard deviation: $\Delta A_\psi = \sqrt{\left\langle \left(A - \langle A \rangle_\psi \right)^2 \right\rangle_\psi} = \sqrt{\langle A^2 \rangle_\psi - \langle A \rangle_\psi^2}$

Cauchy-Schwarz inequality: $|\langle u|v\rangle|^2 \leq \langle u|u\rangle \langle v|v\rangle$

Uncertainty theorem: $\Delta A_\psi \cdot \Delta B_\psi \geq \frac{1}{2} |\langle AB - BA \rangle_\psi|$

Outer Product

$$\begin{aligned} |v\rangle\langle u| &= \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \begin{bmatrix} u_1^* & u_2^* & u_3^* \end{bmatrix} = \begin{bmatrix} v_1 u_1^* & v_1 u_2^* & v_1 u_3^* \\ v_2 u_1^* & v_2 u_2^* & v_2 u_3^* \\ v_3 u_1^* & v_3 u_2^* & v_3 u_3^* \end{bmatrix} \\ |0\rangle\langle 0| &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad |0\rangle\langle 1| = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad |1\rangle\langle 0| = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \\ \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} &= a_{00} |0\rangle\langle 0| + a_{01} |0\rangle\langle 1| + a_{10} |1\rangle\langle 0| + a_{11} |1\rangle\langle 1| \end{aligned}$$

Projection

Projector onto $|u\rangle$: $P_u = |u\rangle\langle u|$

Projection of $|v\rangle$ onto $|u\rangle$: $P_u |v\rangle = |u\rangle\langle u|v\rangle = \langle u|v\rangle |u\rangle$

Projectors are idempotent: $P_u^2 = P_u$

Unitary Matrices

U is unitary: $U^\dagger U = I$ or $U^\dagger = U^{-1}$

The columns (or rows) of a unitary matrix form an orthonormal basis:

$$(U|i\rangle)^\dagger(U|j\rangle) = \langle i|U^\dagger U|j\rangle = \langle i|j\rangle = \delta_{ij}$$

Unitary transformation preserves inner product:

$$\langle U\phi|U\psi\rangle = (U|\phi\rangle)^\dagger(U|\psi\rangle) = \langle\phi|U^\dagger U|\psi\rangle = \langle\phi|\psi\rangle$$

$\{|\phi_i\rangle\}$ are orthonormal and $|\lambda_i| = 1$ \Leftrightarrow $\sum_i \lambda_i |\phi_i\rangle\langle\phi_i|$ is unitary.

If U and V are unitary, so is $U \otimes V$.

$$|\det U| = 1$$

Standard (or Computational) Basis

$$\text{Example: } |0\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad |2\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Orthonormality and completeness: $\langle i|j\rangle = \delta_{ij}$, $\sum_i |i\rangle\langle i| = I$

Vector decomposition: $|v\rangle = \sum_i v_i |i\rangle$, $v_i = \langle i|v\rangle$

Matrix decomposition: $A = \sum_{i,j} a_{ij} |i\rangle\langle j|$, $a_{ij} = \langle i|A|j\rangle$, $A|i\rangle = \sum_j a_{ij} |j\rangle$

Change of Basis

Let $\{|b_i\rangle\}$ and $\{|b'_i\rangle\}$ be complete and orthonormal bases:

$$\langle b_i | b_j \rangle = \langle b'_i | b'_j \rangle = \delta_{ij}, \quad \sum_i |b_i\rangle \langle b_i| = \sum_i |b'_i\rangle \langle b'_i| = I$$

Change of basis from $\{|b_i\rangle\}$ to $\{|b'_i\rangle\}$ via U : $|b'_i\rangle = U |b_i\rangle$

where U is unitary and is given by

$$U = \sum_i |b'_i\rangle \langle b_i| = \sum_{i,j} \langle b_i | b'_j \rangle |b'_i\rangle \langle b'_j| = \begin{bmatrix} \langle b_1 | b'_1 \rangle & \langle b_1 | b'_2 \rangle & \dots \\ \langle b_2 | b'_1 \rangle & \langle b_2 | b'_2 \rangle & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Vectors and Matrices under Change of Basis

$$|v'\rangle = U^\dagger |v\rangle \quad A' = U^\dagger A U$$

$$\langle u' | v' \rangle = \langle u | v \rangle \quad \langle u' | A' | v' \rangle = \langle u | A | v \rangle$$

$$H |\psi_i\rangle = \lambda_i |\psi_i\rangle \rightarrow H' |\psi'_i\rangle = \lambda_i |\psi'_i\rangle \quad \text{tr } A' = \text{tr } A$$

Change from Standard Basis to Eigenvector Basis

Given H is Hermitian and $H |\phi_i\rangle = \lambda_i |\phi_i\rangle$, to change basis from $\{|i\rangle\}$ to $\{|\phi_i\rangle\}$:

$$|\phi_i\rangle = U |i\rangle, \quad U = \sum_i |\phi_i\rangle \langle i|$$

$$H \text{ is diagonal in } \{|\phi_i\rangle\}: H = \sum_i \lambda_i |\phi_i\rangle \langle \phi_i|$$

$$U^\dagger H U \text{ is diagonal in } \{|i\rangle\}: U^\dagger H U = \sum_i \lambda_i |i\rangle \langle i|$$

Tensor Product

$$|u\rangle \otimes |v\rangle \equiv |u\rangle |v\rangle \equiv |uv\rangle$$

$$|uv\rangle = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \otimes \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 |v\rangle \\ u_2 |v\rangle \end{bmatrix} = \begin{bmatrix} u_1 v_1 \\ u_1 v_2 \\ u_2 v_1 \\ u_2 v_2 \end{bmatrix}$$

$$A \otimes B =$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}$$

In general, $(A \otimes B) \neq (B \otimes A)$

$$(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger \quad (A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

$$(A \otimes B)(|u\rangle \otimes |v\rangle) \equiv (A \otimes B)|uv\rangle = |Au\rangle \otimes |Bv\rangle$$

$$A \otimes B = (A \otimes I)(I \otimes B)$$

$$(A_1 \otimes B_1)(A_2 \otimes B_2) \cdots (A_n \otimes B_n) = (A_1 A_2 \cdots A_n) \otimes (B_1 B_2 \cdots B_n)$$

$$(A_1 \otimes A_2 \cdots \otimes A_n)(B_1 \otimes B_2 \cdots \otimes B_n) = (A_1 B_1) \otimes (A_2 B_2) \cdots \otimes (A_n B_n)$$

Tensor product and inner product: $\langle \phi_1 \otimes \phi_2 | \psi_1 \otimes \psi_2 \rangle = \langle \phi_1 | \psi_1 \rangle \langle \phi_2 | \psi_2 \rangle$

Tensor product and outer product:

$$(|i\rangle_1 \otimes |j\rangle_2)(\langle k|_1 \otimes \langle l|_2) = |i\rangle_1 \langle k|_1 \otimes |j\rangle_2 \langle l|_2$$

$$\text{Shorthand notation: } |ij\rangle\langle kl| = |i\rangle\langle k| \otimes |j\rangle\langle l|$$

$$\text{Exponent notation for } \otimes: |x\rangle^{\otimes n} \equiv \bigotimes_{i=1}^n |x\rangle \equiv |xx \cdots x\rangle$$

Basis expansion in linear space \mathbb{C}^{2^n} (i.e., general state vector of an n -qubit system):

$$|\psi\rangle = \sum_{\substack{x_i \in \{0,1\} \\ i \in \{1,2,\dots,n\}}} c_{x_1 x_2 \cdots x_n} |x_1 x_2 \cdots x_n\rangle \equiv \sum_{k=0}^{2^n - 1} c_k |k\rangle$$

Functions of Matrices

If A is Hermitian with real eigenvalues λ_i and eigenvectors $|\phi_i\rangle$, then

$$f(A) = \sum_i f(\lambda_i) |\phi_i\rangle\langle\phi_i|$$

$$e^A = \sum_i e^{\lambda_i} |\phi_i\rangle\langle\phi_i|$$

$$\log A = \sum_i \log \lambda_i |\phi_i\rangle\langle\phi_i| \text{ (for positive } A\text{)}$$

With orthonormal basis $\{|\phi_i\rangle\}$ and $m \in \mathbb{Z}^+$,

$$(|\phi_i\rangle\langle\phi_i|)^m = |\phi_i\rangle\langle\phi_i|, \quad \left(\sum_i \lambda_i |\phi_i\rangle\langle\phi_i| \right)^m = \sum_i \lambda_i^m |\phi_i\rangle\langle\phi_i|$$

$$\text{Alternative definition: } \exp(A) \equiv e^A \equiv \sum_{n=0}^{\infty} \frac{1}{n!} A^n$$

If A is normal, then e^A is also normal, and the eigenvalues of e^A are the exponentials of the eigenvalues of A , with the same eigenvectors.

If H is Hermitian, then e^{iH} is unitary.

If U is unitary, $e^{UAU^\dagger} = Ue^AU^\dagger$

Generalized Euler formula: If γ is real and $A^2 = I$, then $e^{i\gamma A} = \cos \gamma I + i \sin \gamma A$

In general, $e^{A+B} \neq e^A e^B$, unless A and B commute.

BCH formula: $e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]+\dots}$

In general, $e^{A \otimes B} \neq e^A \otimes e^B$, but $e^{A \otimes I + I \otimes B} = e^A \otimes e^B$.

Partial Product Notations

$$U_A |x_1 x_2\rangle \equiv U|x_1\rangle \otimes |x_2\rangle \quad U_B |x_1 x_2\rangle \equiv |x_1\rangle \otimes U|x_2\rangle$$

$$\langle a^{(1)}|x_1 x_2\rangle \equiv \langle a|x_1\rangle |x_2\rangle \quad \langle a^{(2)}|x_1 x_2\rangle \equiv \langle a|x_2\rangle |x_1\rangle$$

$$\begin{aligned} \langle a^{(j)}|x_1 x_2 \cdots x_j \cdots x_n\rangle &\equiv \langle a|x_j\rangle |x_1 x_2 \cdots x_{j-1} x_{j+1} \cdots x_n\rangle \\ &\equiv (I_{j-1} \otimes \langle a| \otimes I_{n-j}) |x_1 x_2 \cdots x_j \cdots x_n\rangle \end{aligned}$$

$$\begin{aligned} \langle a_1^{(1)} a_2^{(2)} \cdots a_m^{(m)}|x_1 x_2 \cdots x_m \cdots x_n\rangle &\equiv \langle a_1 a_2 \cdots a_m|x_1 x_2 \cdots x_m\rangle |x_{m+1} x_{m+2} \cdots x_n\rangle \\ &\equiv (\langle a_1 a_2 \cdots a_m| \otimes I_{n-m}) |x_1 x_2 \cdots x_m \cdots x_n\rangle \end{aligned}$$

$$\text{Given } |\psi\rangle = \sum_{x_1, x_2, \dots, x_n \in \{0,1\}} c_{x_1 x_2 \cdots x_n} |x_1 x_2 \cdots x_n\rangle,$$

$$\langle a_1^{(1)} a_2^{(2)} \cdots a_m^{(m)}|\psi\rangle = \sum_{x_{m+1}, \dots, x_n \in \{0,1\}} c_{a_1 a_2 \cdots a_m x_{m+1} \cdots x_n} |x_{m+1} \cdots x_n\rangle$$

Trace

$$\text{tr } A \equiv \text{tr}(A) \equiv \sum_{i=1}^n a_{ii}$$

$$\text{tr}(A^T) = \text{tr}(A) \quad \text{tr}(A^\dagger) = \text{tr}(A^*) = (\text{tr } A)^*$$

$$\text{tr}(cA) = c \text{tr } A \quad \text{tr}(A + B) = \text{tr } A + \text{tr } B$$

$$\text{tr}(AB) = \text{tr}(BA) \quad (\text{Note in general, } AB \neq BA, \text{ and } \text{tr}(AB) \neq \text{tr}(A) \text{tr}(B))$$

$$\text{tr}([A, B]) = 0$$

For Pauli matrices: $\text{tr } \sigma_j = 0$, $\text{tr}(\sigma_j \sigma_k) = 2\delta_{jk}$

Similarity invariance: $\text{tr}(PAP^{-1}) = \text{tr } A$

As sum of eigenvalues: $\text{tr}(A) = \sum_{i=1}^n \lambda_i, \quad \text{tr}(A^k) = \sum_{i=1}^n \lambda_i^k$

Tensor product property: $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$

Outer product property: $\langle v|u \rangle = \text{tr}(|u\rangle\langle v|), \quad \langle v|A|u \rangle = \text{tr}(A|u\rangle\langle v|)$

Cyclic product property: $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB) \quad (\neq \text{tr}(BAC))$

For orthonormal basis $\{|\phi_i\rangle\}$, $\text{tr}\left(\sum_{i=1}^n |\phi_i\rangle\langle\phi_i|\right) = n, \quad \text{tr } A = \sum_{i=1}^n \langle\phi_i|A|\phi_i\rangle$

A useful identity: $\langle\phi_i|A|\phi_i\rangle = \text{tr}(\langle\phi_i|A|\phi_i\rangle) = \text{tr}(A|\phi_i\rangle\langle\phi_i|)$

Cauchy-Schwarz inequality: $0 \leq [\text{tr}(AB)]^2 \leq \text{tr}(A^2) \text{tr}(B^2) \leq [\text{tr}(A)]^2 [\text{tr}(B)]^2$

For normal matrix: $\text{tr}(A) = \log(\det(e^A))$

Determinant

$$\det(A^T) = \det(A) \quad \det(A^\dagger) = \det(A^*) = \det(A)^*$$

$$\det(A^{-1}) = \det(A)^{-1}$$

$$\det(AB) = \det(A)\det(B)$$

$$\det(A \otimes B) = (\det A)^n (\det B)^m \quad (\text{where } m \text{ is the dimension of } A, \text{ and } n \text{ of } B.)$$

$$\text{For unitary matrix: } |\det(U)| = 1$$

$$\text{For normal matrix: } \det(A) = \prod_i \lambda_i \quad \det(e^A) = e^{\text{tr } A}$$

Vector Space of Matrices

Inner product of two matrices:

$$\langle A, B \rangle \equiv A \cdot B \equiv \sum_{i=1}^m \sum_{j=1}^n a_{ij}^* b_{ij} = \text{tr}(A^\dagger B) = \text{tr}(BA^\dagger)$$

Decomposition of $A \in \mathbb{C}^{2^n \times 2^n}$ in Pauli basis:

$$P_i \in \{I, X, Y, Z\}^{\otimes n}, \quad P_i \cdot P_j = 2^n \delta_{ij}, \quad \text{with } i, j = 1, 2, \dots, 4^n$$

$$A = \sum_{i=1}^{4^n} a_i P_i, \quad a_i = \frac{1}{2^n} P_i \cdot A \equiv \frac{1}{2^n} \text{tr}(AP_i^\dagger) = \frac{1}{2^n} \text{tr}(AP_i)$$

D Pauli Matrices

Definitions

$$\begin{aligned}\sigma_0 &= I = |0\rangle\langle 0| + |1\rangle\langle 1| &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \sigma_1 = \sigma_x &= X = |0\rangle\langle 1| + |1\rangle\langle 0| &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ \sigma_2 = \sigma_y &= Y = |0\rangle\langle 1| (-i) + |1\rangle\langle 0| i &= \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \\ \sigma_3 = \sigma_z &= Z = |0\rangle\langle 0| - |1\rangle\langle 1| &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\end{aligned}$$

Eigenvalues and Vectors

Pauli Matrix	Eigenvalue	Eigenvector
$\sigma_z \equiv Z$	1	$ 0\rangle$
	-1	$ 1\rangle$
$\sigma_x \equiv X$	1	$ +\rangle = \frac{1}{\sqrt{2}}(0\rangle + 1\rangle)$
	-1	$ -\rangle = \frac{1}{\sqrt{2}}(0\rangle - 1\rangle)$
$\sigma_y \equiv Y$	1	$ +_i\rangle = \frac{1}{\sqrt{2}}(0\rangle + i 1\rangle)$
	-1	$ -_i\rangle = \frac{1}{\sqrt{2}}(0\rangle - i 1\rangle)$

Properties

For $i, j, k \in \{1, 2, 3\}$,

Unitary and Hermitian: $\sigma_j^2 = \sigma_0$, $\sigma_j^\dagger = \sigma_j$

Commutation relation: $[\sigma_j, \sigma_k] = 2i\varepsilon_{jkl}\sigma_l$, or $\sigma_1\sigma_2 = \sigma_3$, etc.

Anti-commutation relation: $\{\sigma_j, \sigma_k\} = 2\delta_{jk}\sigma_0$, or $\sigma_j\sigma_k = -\sigma_k\sigma_j$ for $j \neq k$

Orthogonality: $\sigma_j \cdot \sigma_k \equiv \text{tr}(\sigma_j^\dagger \sigma_k) = \text{tr}(\sigma_j \sigma_k) = 2\delta_{jk}$

In addition, $\text{tr } \sigma_j = 0$, $\det \sigma_j = -1$, $\sigma_j \sigma_k = i\varepsilon_{jkl}\sigma_l + \delta_{jk}\sigma_0$

Basis for 2×2 Matrices

Any matrix $A \in \mathbb{C}^{2 \times 2}$ can be expanded over the set $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$:

$$A = \sum_{j=0}^3 a_j \sigma_j, \text{ where } a_0 = \frac{1}{2} \text{tr } A, \text{ and } a_j = \frac{1}{2} \text{tr}(\sigma_j A).$$

Representing Spins in Any Direction

For any real unit vector $\hat{u} = (u_x, u_y, u_z)$ where $u_x^2 + u_y^2 + u_z^2 = 1$,

define $\sigma_u = \hat{u} \cdot \sigma = u_x \sigma_x + u_y \sigma_y + u_z \sigma_z$.

For $\hat{u} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$,

$$\sigma_u = \sin \theta \cos \phi \sigma_x + \sin \theta \sin \phi \sigma_y + \cos \theta \sigma_z = \begin{bmatrix} \cos \theta & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta \end{bmatrix}$$

Eigenvectors:

$$\begin{aligned} |+u\rangle &= \cos \frac{\theta}{2} |0\rangle + \sin \frac{\theta}{2} e^{i\phi} |1\rangle = \begin{bmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\phi} \end{bmatrix} \\ |-u\rangle &= -\sin \frac{\theta}{2} |0\rangle + \cos \frac{\theta}{2} e^{i\phi} |1\rangle = \begin{bmatrix} -\sin \frac{\theta}{2} \\ \cos \frac{\theta}{2} e^{i\phi} \end{bmatrix} \end{aligned}$$

Expected values:

$$\langle +u | \sigma_x | +u \rangle = \sin \theta \cos \phi$$

$$\langle +u | \sigma_y | +u \rangle = \sin \theta \sin \phi$$

$$\langle +u | \sigma_z | +u \rangle = \cos \theta$$

Exponentiation

For any real γ , $e^{i\gamma\sigma_j} = \cos \gamma \sigma_0 + i \sin \gamma \sigma_j$

Also, $e^{i\gamma\sigma_u} = \cos \gamma \sigma_0 + i \sin \gamma \sigma_u$

Rotation Operators

$$R_j(\gamma) \equiv e^{-i\frac{\gamma}{2}\sigma_j} = \sigma_0 \cos \frac{\gamma}{2} - i\sigma_j \sin \frac{\gamma}{2}$$

Explicitly,

$$R_1(\gamma) \equiv R_x(\gamma) = \begin{bmatrix} \cos \frac{\gamma}{2} & -i \sin \frac{\gamma}{2} \\ -i \sin \frac{\gamma}{2} & \cos \frac{\gamma}{2} \end{bmatrix}$$

$$R_2(\gamma) \equiv R_y(\gamma) = \begin{bmatrix} \cos \frac{\gamma}{2} & -\sin \frac{\gamma}{2} \\ \sin \frac{\gamma}{2} & \cos \frac{\gamma}{2} \end{bmatrix}$$

$$R_3(\gamma) \equiv R_z(\gamma) = \begin{bmatrix} e^{-i\frac{\gamma}{2}} & 0 \\ 0 & e^{i\frac{\gamma}{2}} \end{bmatrix}$$

Rotation about axis $\hat{u} = (n_x, n_y, n_z)$:

$$R_u(\gamma) = e^{-i\frac{\gamma}{2}\sigma_u} = \sigma_0 \cos \frac{\gamma}{2} - i\sigma_u \sin \frac{\gamma}{2}$$

Rotation from $|0\rangle$ to $|+u\rangle$ and $| -u\rangle$:

$$\begin{aligned}|+u\rangle &= R_z(\phi)R_y(\theta)|0\rangle \\ | -u\rangle &= R_z(\phi)R_y(-\theta)|0\rangle\end{aligned}$$

Other Forms

$$\begin{aligned}\sigma_+ &\equiv \frac{1}{2}(\sigma_1 + i\sigma_2) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} & \sigma_- &\equiv \frac{1}{2}(\sigma_1 - i\sigma_2) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \\ \sigma_h &\equiv \frac{1}{2}(\sigma_0 + \sigma_3) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} & \sigma_v &\equiv \frac{1}{2}(\sigma_0 - \sigma_3) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\end{aligned}$$

Gate Relations

$$XY = -YX = iZ$$

$$YZ = -ZY = iX$$

$$ZX = -XZ = iY$$

$$XXX = X, \quad XYX = -Y, \quad XZX = -Z,$$

$$YXY = -X, \quad YYY = Y, \quad YZY = -Z,$$

$$ZXZ = -X, \quad ZYZ = -Y, \quad ZZZ = Z$$

$$HXH = Z, \text{ where } H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$HYH = -Y$$

$$HZH = X$$

$$SXS^\dagger = Y, \text{ where } S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$$

$$SYS^\dagger = -X$$

$$SZS^\dagger = Z$$

Pauli Strings

Definition: $P = A_1 \otimes \cdots \otimes A_n$ with $A_i \in \{I, X, Y, Z\}$, or $P \in \{I, X, Y, Z\}^{\otimes n}$

Orthogonality: $P_i \cdot P_j \equiv \text{tr}(P_i^\dagger P_j) = 2^n \delta_{ij}$, with $i, j = 1, 2, \dots, 4^n$.

Two Pauli strings P_i and P_j commute if they have an even number of qubits where their corresponding Pauli matrices are different and anticommute (e.g., one is X and the other is Y or Z , but not I).

Bibliography

Books

- [1] Claude Cohen-Tannoudji, Bernard Diu, and Franck Laloë. *Quantum Mechanics, Vol 1: Basic Concepts, Tools, and Applications*. Wiley, 2019. ISBN: 978-3-527-34553-3 (cited on page 35).
- [2] Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *Quantum Mechanics (Feynman Lectures on Physics, Volume 3)*. Basic Books, Oct. 4, 2011. ISBN: 978-0465025015 (cited on page 36).
- [3] David J. Griffiths and Darrell F. Schroeter. *Introduction to Quantum Mechanics*. English. 3rd edition. Cambridge University Press, Aug. 2018, page 508. ISBN: 978-1107189638 (cited on page 390).
- [4] Kurt Jacobs. *Quantum Measurement Theory and its Applications*. English. 1st. Cambridge University Press, 2014, page 554. ISBN: 978-1-107-02548-6 (cited on page 77).
- [5] Jonathan Katz and Yehuda Lindell. *Introduction to Modern Cryptography*. CRC Press, 2014. ISBN: 978-1-4665-7027-6 (cited on page 130).
- [6] Junichiro Kono. *Quantum Mechanics for Tomorrow's Engineers: New Edition*. English. New edition. Cambridge University Press, Sept. 29, 2022, page 350. ISBN: 978-1108842587 (cited on page 35).
- [7] Michael A Nielsen and Isaac L Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. ISBN: 978-1-107-00217-3 (cited on pages 180, 213, 259, 435).
- [8] J. J. Sakurai and Jim Napolitano. *Modern Quantum Mechanics*. Cambridge University Press, 2017. ISBN: 978-1-108-47322-4 (cited on pages 36, 86, 99).
- [9] Gilbert Strang. *Introduction to Linear Algebra*. English. 6th edition. Wellesley-Cambridge Press, Apr. 2023, page 440. ISBN: 978-1733146678 (cited on page 18).
- [10] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. ISBN: 978-1-107-18056-7 (cited on page 435).
- [11] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2017. ISBN: 978-1-107-17616-4 (cited on pages 77, 435).

Articles

- [12] Scott Aaronson and Paul Christiano. Quantum Money from Hidden Subspaces. In: (2012) (cited on page 326).
- [13] R. Alléaume et al. Using quantum key distribution for cryptographic purposes: A survey. In: Theoretical Computer Science 560 (2014), pages 62–81. ISSN: 0304-3975. DOI: [10.1016/j.tcs.2014.09.018](https://doi.org/10.1016/j.tcs.2014.09.018) (cited on pages 130, 282).

- [14] Rameez Asif. Post-Quantum Cryptosystems for Internet-of-Things: A Survey on Lattice-Based Algorithms. In: IoT 2.1 (2021), pages 71–91. ISSN: 2624-831X. DOI: [10.3390/iot2010005](https://doi.org/10.3390/iot2010005) (cited on pages 125, 130, 327).
- [15] Alain Aspect, Jean Dalibard, and G'erard Roger. Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment: A new violation of Bell's inequalities. In: Physical Review Letters 49.2 (1982), page 91. DOI: [10.1103/PhysRevLett.49.91](https://doi.org/10.1103/PhysRevLett.49.91) (cited on page 240).
- [16] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental test of Bell's inequalities using time-varying analyzers. In: Physical Review Letters 49.25 (1982), pages 1804–1807. DOI: [10.1103/PhysRevLett.49.1804](https://doi.org/10.1103/PhysRevLett.49.1804) (cited on page 231).
- [17] Alain Aspect, Philippe Grangier, and Gérard Roger. Experimental tests of realistic local theories via Bell's theorem. In: Physical Review Letters 47.7 (1981), pages 460–463. DOI: [10.1103/PhysRevLett.47.460](https://doi.org/10.1103/PhysRevLett.47.460) (cited on page 231).
- [18] Adriano Barenco et al. Elementary gates for quantum computation. In: Physical Review A 52.5 (1995), page 3457. DOI: [10.1103/PhysRevA.52.3457](https://doi.org/10.1103/PhysRevA.52.3457) (cited on page 180).
- [19] Elisa Bäumer et al. Efficient Long-Range Entanglement using Dynamic Circuits. In: (2023) (cited on page 335).
- [20] John S Bell. On the Einstein Podolsky Rosen Paradox. In: Physics Physique Fizika 1 (3 Nov. 1964), pages 195–200. DOI: [10.1103/PhysicsPhysiqueFizika.1.195](https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195) (cited on page 240).
- [21] Charles H Bennett and Gilles Brassard. Quantum Cryptography: Public Key Distribution and Coin Tossing. In: Proceedings of IEEE International Conference on Computers, Systems and Signal Processing (1984), pages 175–179. DOI: [10.1016/j.tcs.2014.05.025](https://doi.org/10.1016/j.tcs.2014.05.025) (cited on pages 130, 282).
- [22] Charles H. Bennett. Quantum Cryptography Using Any Two Nonorthogonal States. In: Physical Review Letters 68.21 (1992), page 3121. DOI: [10.1103/PhysRevLett.68.3121](https://doi.org/10.1103/PhysRevLett.68.3121) (cited on page 130).
- [23] Kishor Bharti et al. Noisy intermediate-scale quantum algorithms. In: Reviews of Modern Physics 94.1 (Feb. 2022). DOI: [10.1103/revmodphys.94.015004](https://doi.org/10.1103/revmodphys.94.015004) (cited on page 299).
- [24] A. Bilyk, J. Doliskani, and Z. Gong. Cryptanalysis of Three Quantum Money Schemes. In: Quantum Information Processing 22 (2023), page 177. DOI: [10.1007/s11128-023-03919-0](https://doi.org/10.1007/s11128-023-03919-0) (cited on page 326).
- [25] Jan Bouda and Vladimír Bužek. Entanglement swapping between multi-qudit systems. In: Journal of Physics A: Mathematical and General 34.20 (May 2001), page 4301. DOI: [10.1088/0305-4470/34/20/304](https://doi.org/10.1088/0305-4470/34/20/304) (cited on page 264).
- [26] Dik Bouwmeester et al. Experimental quantum teleportation. In: Nature 390.6660 (Dec. 1997), pages 575–579. DOI: [10.1038/37539](https://doi.org/10.1038/37539) (cited on page 259).
- [27] Dik Bouwmeester et al. Observation of three-photon Greenberger-Horne-Zeilinger entanglement. In: Physical review letters 82.7 (1999), page 1345. DOI: [10.1103/PhysRevLett.82.1345](https://doi.org/10.1103/PhysRevLett.82.1345) (cited on page 213).
- [28] Sergey Bravyi, David Gosset, and Yinchen Liu. How to Simulate Quantum Measurement without Computing Marginals. In: Phys. Rev. Lett. 128 (22 June 2022), page 220503. DOI: [10.1103/PhysRevLett.128.220503](https://doi.org/10.1103/PhysRevLett.128.220503) (cited on page 389).

- [29] Sergey Bravyi et al. High-threshold and low-overhead fault-tolerant quantum memory. In: (2023) (cited on page 389).
- [30] H.-J. Briegel et al. Quantum Repeaters: The Role of Imperfect Local Operations in Quantum Communication. In: Phys. Rev. Lett. 81 (26 Dec. 1998), pages 5932–5935. DOI: [10.1103/PhysRevLett.81.5932](https://doi.org/10.1103/PhysRevLett.81.5932) (cited on page 264).
- [31] Aharon Brodutch et al. An adaptive attack on Wiesner’s quantum money. In: (2016) (cited on page 321).
- [32] D. Bruß et al. Distributed Quantum Dense Coding. In: Phys. Rev. Lett. 93 (21 Nov. 2004), page 210501. DOI: [10.1103/PhysRevLett.93.210501](https://doi.org/10.1103/PhysRevLett.93.210501) (cited on page 250).
- [33] Zhenyu Cai, Xiaosi Xu, and Simon C. Benjamin. Mitigating coherent noise using Pauli conjugation. In: npj Quantum Information 6.1 (2020), page 17. ISSN: 2056-6387. DOI: [10.1038/s41534-019-0233-0](https://doi.org/10.1038/s41534-019-0233-0) (cited on page 389).
- [34] Avimita Chatterjee, Koustubh Phalak, and Swaroop Ghosh. Quantum Error Correction For Dummies. In: (2023) (cited on pages 388, 389).
- [35] Edward H. Chen et al. Realizing the Nishimori transition across the error threshold for constant-depth quantum circuits. In: (2023) (cited on page 335).
- [36] Kevin S. Chou et al. Deterministic teleportation of a quantum gate between two logical qubits. In: Nature 561.7723 (2018), pages 368–373. ISSN: 1476-4687. DOI: [10.1038/s41586-018-0470-y](https://doi.org/10.1038/s41586-018-0470-y) (cited on page 277).
- [37] A. Dan et al. Clustering approach for solving traveling salesman problems via Ising model based solver. In: (2020), pages 1–6. DOI: [10.1109/DAC18072.2020.9218695](https://doi.org/10.1109/DAC18072.2020.9218695) (cited on page 312).
- [38] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? In: Physical Review 47.10 (1935), page 777. DOI: [10.1103/PhysRev.47.777](https://doi.org/10.1103/PhysRev.47.777) (cited on page 240).
- [39] Artur K Ekert. Quantum cryptography based on Bell’s theorem. In: Physical Review Letters 67.6 (1991), page 661. DOI: [10.1103/PhysRevLett.67.661](https://doi.org/10.1103/PhysRevLett.67.661) (cited on pages 130, 281).
- [40] Avshalom C. Elitzur and Lev Vaidman. Quantum mechanical interaction-free measurements. In: Foundations of Physics 23.7 (July 1993), pages 987–997. DOI: [10.1007/bf00736012](https://doi.org/10.1007/bf00736012) (cited on page 315).
- [41] Edward Farhi et al. Quantum money from knots. In: (2010) (cited on page 326).
- [42] Arman Rasoodl Faridi et al. Blockchain in the Quantum World. In: International Journal of Advanced Computer Science and Applications 13.1 (2022). DOI: [10.14569/ijacsa.2022.0130167](https://doi.org/10.14569/ijacsa.2022.0130167) (cited on page 327).
- [43] Roland C. Farrell et al. Scalable Circuits for Preparing Ground States on Digital Quantum Computers: The Schwinger Model Vacuum on 100 Qubits. In: (2023) (cited on page 335).
- [44] A. Furusawa et al. Unconditional Quantum Teleportation. In: Science 282.5389 (1998), pages 706–709. DOI: [10.1126/science.282.5389.706](https://doi.org/10.1126/science.282.5389.706) (cited on page 259).
- [45] Marissa Giustina et al. A significant-loophole-free test of Bell’s theorem with entangled photons. In: Physical Review Letters 115.25 (2015), page 250401. DOI: [10.1103/PhysRevLett.115.250401](https://doi.org/10.1103/PhysRevLett.115.250401) (cited on page 232).

- [46] Daniel Gottesman and Isaac L. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. In: *Nature* 402.6760 (Nov. 1999), pages 390–393. DOI: [10.1038/46503](https://doi.org/10.1038/46503) (cited on page 276).
- [47] Trent M. Graham et al. Superdense teleportation using hyperentangled photons. In: *Nature Communications* 6.1 (May 2015), page 7185. ISSN: 2041-1723. DOI: [10.1038/ncomms8185](https://doi.org/10.1038/ncomms8185) (cited on page 259).
- [48] Daniel M Greenberger, Michael A Horne, and Anton Zeilinger. Going beyond Bell's theorem. In: (1989), pages 69–72. DOI: [10.1007/978-94-017-0849-4_10](https://doi.org/10.1007/978-94-017-0849-4_10) (cited on page 213).
- [49] Shouzhen Gu, Alex Retzker, and Aleksander Kubica. Fault-tolerant quantum architectures based on erasure qubits. In: (2023) (cited on page 389).
- [50] Yu Guo et al. Advances in Quantum Dense Coding. In: *Advanced Quantum Technologies* 2.5-6 (2019), page 1900011. DOI: [10.1002/qute.201900011](https://doi.org/10.1002/qute.201900011) (cited on page 250).
- [51] Mauricio Gutiérrez and Kenneth R. Brown. Comparison of a quantum error-correction threshold for exact and approximate errors. In: *Physical Review A* 91.2 (Feb. 2015). DOI: [10.1103/physreva.91.022335](https://doi.org/10.1103/physreva.91.022335) (cited on page 389).
- [52] Laszlo Gyongyosi and Sandor Imre. A Survey on quantum computing technology. In: *Computer Science Review* 31 (2019), pages 51–71. ISSN: 1574-0137. DOI: [10.1016/j.cosrev.2018.11.002](https://doi.org/10.1016/j.cosrev.2018.11.002) (cited on page 57).
- [53] Johannes Handsteiner et al. Cosmic Bell Test: Measurement Settings from Milky Way Stars. In: *Physical Review Letters* 118.6 (2017). DOI: [10.1103/PhysRevLett.118.060401](https://doi.org/10.1103/PhysRevLett.118.060401) (cited on page 241).
- [54] B Hensen et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. In: *Nature* 526.7575 (2015), pages 682–686. DOI: [10.1038/nature15759](https://doi.org/10.1038/nature15759) (cited on pages 232, 240).
- [55] Ryszard Horodecki et al. Quantum entanglement. In: *Reviews of Modern Physics* 81.2 (2009), pages 865–942. DOI: [10.1103/RevModPhys.81.865](https://doi.org/10.1103/RevModPhys.81.865) (cited on page 240).
- [56] Yun-Feng Huang et al. Experimental Teleportation of a Quantum Controlled-NOT Gate. In: *Phys. Rev. Lett.* 93 (24 Dec. 2004), page 240501. DOI: [10.1103/PhysRevLett.93.240501](https://doi.org/10.1103/PhysRevLett.93.240501) (cited on page 276).
- [57] Daniel M. Kane, Shahed Sharif, and Alice Silverberg. Quantum Money from Quaternion Algebras. In: (2022) (cited on page 326).
- [58] Daniel Keith et al. Ramped measurement technique for robust high-fidelity spin qubit readout. In: *Science Advances* 8.36 (2022), eabq0455. DOI: [10.1126/sciadv.abq0455](https://doi.org/10.1126/sciadv.abq0455) (cited on page 53).
- [59] Andrey Boris Khesin, Jonathan Z. Lu, and Peter W. Shor. Publicly verifiable quantum money from random lattices. In: (2022) (cited on page 326).
- [60] Dai-Gyoung Kim et al. Enhanced quantum teleportation using multi-qubit logical states. In: *Results in Physics* 50 (2023), page 106565. ISSN: 2211-3797. DOI: [10.1016/j.rinp.2023.106565](https://doi.org/10.1016/j.rinp.2023.106565) (cited on page 277).
- [61] Sangbae Kim and Byoung S. Ham. Observations of the delayed-choice quantum eraser using coherent photons. In: *Scientific Reports* 13.1 (June 16, 2023), page 9758. ISSN: 2045-2322. DOI: [10.1038/s41598-023-36590-7](https://doi.org/10.1038/s41598-023-36590-7) (cited on page 30).

- [62] Youngseok Kim et al. Evidence for the utility of quantum computing before fault tolerance. In: *Nature* 618.7965 (June 2023), pages 500–505. ISSN: 1476-4687. DOI: [10.1038/s41586-023-06096-3](https://doi.org/10.1038/s41586-023-06096-3) (cited on page 335).
- [63] Brian T. Kirby et al. Entanglement swapping of two arbitrarily degraded entangled states. In: *Phys. Rev. A* 94 (1 July 2016), page 012336. DOI: [10.1103/PhysRevA.94.012336](https://doi.org/10.1103/PhysRevA.94.012336) (cited on page 264).
- [64] Raymond Laflamme et al. Perfect Quantum Error Correction Code. In: (1996) (cited on page 388).
- [65] Jan-Åke Larsson. Loopholes in Bell inequality tests of local realism. In: *Journal of Physics A: Mathematical and Theoretical* 47.42 (2014). DOI: [10.1088/1751-8113/47/42/424003](https://doi.org/10.1088/1751-8113/47/42/424003) (cited on page 240).
- [66] Göran Lindblad. Completely positive maps and entropy inequalities. In: *Communications in Mathematical Physics* 40.2 (1975), pages 147–151. DOI: [10.1007/BF01609396](https://doi.org/10.1007/BF01609396) (cited on page 429).
- [67] J. Liu, H. Montgomery, and M. Zhandry. Another Round of Breaking and Making Quantum Money. In: *Lecture Notes in Computer Science* 14004 (2023). Edited by C. Hazay and M. Stam. DOI: [10.1007/978-3-031-30545-0_21](https://doi.org/10.1007/978-3-031-30545-0_21) (cited on page 326).
- [68] G. Lugilde Fernández, E.F. Combarro, and I.F. Rúa. Quantum measurement detection algorithms. In: *Quantum Information Process* 21.274 (2022). DOI: [10.1007/s11128-022-03614-6](https://doi.org/10.1007/s11128-022-03614-6) (cited on page 318).
- [69] Andrew Lutomirski. An online attack against Wiesner’s quantum money. In: (2010) (cited on page 320).
- [70] Chetan Nayak et al. Non-Abelian anyons and topological quantum computation. In: *Rev. Mod. Phys.* 80 (3 Sept. 2008), pages 1083–1159. DOI: [10.1103/RevModPhys.80.1083](https://doi.org/10.1103/RevModPhys.80.1083) (cited on page 391).
- [71] A. K. Pati, P. Parashar, and P. Agrawal. Probabilistic superdense coding. In: *Phys. Rev. A* 72 (1 July 2005), page 012329. DOI: [10.1103/PhysRevA.72.012329](https://doi.org/10.1103/PhysRevA.72.012329) (cited on page 250).
- [72] Asher Peres. All the Bell Inequalities. In: (1999) (cited on page 240).
- [73] Christophe Piveteau et al. Error Mitigation for Universal Gates on Encoded Qubits. In: *Phys. Rev. Lett.* 127 (20 Nov. 2021), page 200505. DOI: [10.1103/PhysRevLett.127.200505](https://doi.org/10.1103/PhysRevLett.127.200505) (cited on page 389).
- [74] Lukas Postler et al. Demonstration of fault-tolerant universal quantum gate operations. In: *Nature* 605.7911 (2022), pages 675–680. ISSN: 1476-4687. DOI: [10.1038/s41586-022-04721-1](https://doi.org/10.1038/s41586-022-04721-1) (cited on page 277).
- [75] Gustavo Rigolin. Quantum teleportation of an arbitrary two-qubit state and its relation to multipartite entanglement. In: *Physical Review A* 71.3 (Mar. 2005). DOI: [10.1103/physreva.71.032303](https://doi.org/10.1103/physreva.71.032303) (cited on page 259).
- [76] A. Rubenok et al. Real-world two-photon interference and proof-of-principle quantum key distribution immune to detector attacks. In: *Physical Review Letters* 111.13 (2013), page 130501. DOI: [10.1103/PhysRevLett.111.130501](https://doi.org/10.1103/PhysRevLett.111.130501) (cited on page 282).
- [77] Maria E. Sabani et al. Evaluation and Comparison of Lattice-Based Cryptosystems for a Secure Quantum Computing Era. In: *Electronics* 12.12 (2023). ISSN: 2079-9292. DOI: [10.3390/electronics12122643](https://doi.org/10.3390/electronics12122643) (cited on page 327).

- [78] M. Sasaki et al. Field test of quantum key distribution in the Tokyo QKD Network. In: *Optics Express* 19.11 (2011), pages 10387–10409. doi: [10.1364/OE.19.010387](https://doi.org/10.1364/OE.19.010387) (cited on page 282).
- [79] Valerio Scarani et al. Quantum Cryptography Protocols Robust Against Photon Number Splitting Attacks. In: *Physical Review Letters* 92.5 (2004), page 057901. doi: [10.1103/PhysRevLett.92.057901](https://doi.org/10.1103/PhysRevLett.92.057901) (cited on page 130).
- [80] Lynden K Shalm et al. A strong loophole-free test of local realism. In: *Physical Review Letters* 115.25 (2015), page 250402. doi: [10.1103/PhysRevLett.115.250402](https://doi.org/10.1103/PhysRevLett.115.250402) (cited on page 232).
- [81] Oles Shtanko et al. Uncovering Local Integrability in Quantum Many-Body Dynamics. In: (2023) (cited on page 335).
- [82] Mitali Sisodia et al. Design and experimental realization of an optimal scheme for teleportation of an n-qubit quantum state. In: *Quantum Information Processing* 16.12 (2017), page 292. ISSN: 1573-1332. doi: [10.1007/s11128-017-1744-2](https://doi.org/10.1007/s11128-017-1744-2) (cited on page 276).
- [83] Erik Sjöqvist. A new phase in quantum computation. In: *Physics* 1 (Nov. 2008), page id. 35. doi: [10.1103/Physics.1.35](https://doi.org/10.1103/Physics.1.35) (cited on page 390).
- [84] A. M. Steane. Simple quantum error-correcting codes. In: *Physical Review A* 54.6 (1996), page 4741. doi: [10.1103/physreva.54.4741](https://doi.org/10.1103/physreva.54.4741) (cited on page 388).
- [85] Ady Stern and Netanel H. Lindner. Topological Quantum Computation—From Basic Concepts to First Experiments. In: *Science* 339.6124 (2013), pages 1179–1184. doi: [10.1126/science.1231473](https://doi.org/10.1126/science.1231473) (cited on page 391).
- [86] Damien Stucki et al. Continuous high speed coherent one-way quantum key distribution. In: *Optics Express* 17.16 (July 2009), page 13326. doi: [10.1364/oe.17.013326](https://doi.org/10.1364/oe.17.013326) (cited on page 130).
- [87] Miroslav Urbanek et al. Mitigating Depolarizing Noise on Quantum Computers with Noise-Estimation Circuits. In: *Phys. Rev. Lett.* 127 (27 Dec. 2021), page 270502. doi: [10.1103/PhysRevLett.127.270502](https://doi.org/10.1103/PhysRevLett.127.270502) (cited on page 390).
- [88] Andrew Wack et al. Quality, Speed, and Scale: three key attributes to measure the performance of near-term quantum computers. In: (2021) (cited on page 389).
- [89] Gregor Weihs et al. Violation of Bell’s inequality under strict Einstein locality conditions. In: *Physical Review Letters* 81.23 (1998), pages 5039–5043. doi: [10.1103/PhysRevLett.81.5039](https://doi.org/10.1103/PhysRevLett.81.5039) (cited on page 231).
- [90] Stephen Wiesner. Conjugate Coding. In: *SIGACT News* 15.1 (Jan. 1983), pages 78–88. ISSN: 0163-5700. doi: [10.1145/1008908.1008920](https://doi.org/10.1145/1008908.1008920) (cited on page 319).
- [91] Yue Wu et al. Erasure conversion for fault-tolerant quantum computing in alkali-earth Rydberg atom arrays. In: *Nature Communications* 13.1 (Aug. 2022). ISSN: 2041-1723. doi: [10.1038/s41467-022-32094-6](https://doi.org/10.1038/s41467-022-32094-6) (cited on page 389).
- [92] Toshiki Yasuda et al. Quantum reservoir computing with repeated measurements on superconducting devices. In: (2023) (cited on page 335).
- [93] Hongye Yu, Yusheng Zhao, and Tzu-Chieh Wei. Simulating large-size quantum spin chains on cloud-based superconducting quantum computers. In: *Physical Review Research* 5.1 (Mar. 2023). ISSN: 2643-1564. doi: [10.1103/physrevres.5.013183](https://doi.org/10.1103/physrevres.5.013183) (cited on page 335).

-
- [94] M. Żukowski et al. ‘Event-ready-detectors’ Bell experiment via entanglement swapping. In: Phys. Rev. Lett. 71 (26 Dec. 1993), pages 4287–4290. DOI: [10.1103/PhysRevLett.71.4287](https://doi.org/10.1103/PhysRevLett.71.4287) (cited on page 264).
 - [95] Sultan M. Zangi et al. Entanglement Swapping and Swapped Entanglement. In: Entropy 25.3 (Feb. 2023), page 415. ISSN: 1099-4300. DOI: [10.3390/e25030415](https://doi.org/10.3390/e25030415) (cited on page 264).
 - [96] Mark Zhandry. Quantum Lightning Never Strikes the Same State Twice. In: (2017) (cited on page 326).
 - [97] Jiang Zhang et al. Geometric and holonomic quantum computation. In: Physics Reports 1027 (2023), pages 1–53. ISSN: 0370-1573. DOI: [10.1016/j.physrep.2023.07.004](https://doi.org/10.1016/j.physrep.2023.07.004) (cited on page 390).
 - [98] Tingting Zhang and Jie Han. Efficient Traveling Salesman Problem Solvers Using the Ising Model with Simulated Bifurcation. In: DATE ’22 (2022), pages 548–551. DOI: [10.23919/DATEx54114.2022.9774576](https://doi.org/10.23919/DATEx54114.2022.9774576) (cited on page 312).

List of Figures

1.1	Role of Quantum Mechanics in Quantum Computing	4
1.2	Light Polarization in Classical and Quantum Realms	7
1.3	A Light Polarization Experiment	9
1.4	Relationship between H/V and D/A Polarizations	12
1.5	Elliptical Polarization	15
1.6	Illustration of the Quantum Observable Postulate	18
1.7	Measuring Photons Through a Polarizer	25
1.8	Repeated Identical Measurements	26
1.9	Measuring Photons Through Crossed Polarizers	27
1.10	Measuring Photons: Malus' Law	29
1.11	Mach-Zehnder Interferometer	31
1.12	Double Slit Interference Experiment	32
1.13	Summary of the Observable and Measurement Postulates	38
2.1	1/2 Spin	43
2.2	Spherical Coordinate System	46
2.3	Key Points on the Bloch Sphere	48
2.4	Pauli X , Y , and Z as Rotations on the Bloch Sphere	50
2.5	Stern-Gerlach Experiment	51
2.6	Cascaded Stern-Gerlach Experiments	52
3.1	Visualization of qubit states on the Bloch Sphere	61
3.2	Basis Rotation $\{ 0\rangle, 1\rangle\} \rightarrow \{ +\rangle, -\rangle\}$	64
3.3	Illustration of Basis Rotation $\{ b_i\rangle\} \rightarrow \{ b'_i\rangle\}$	65
3.4	Measurements in $\{ 0\rangle, 1\rangle\}$ and $\{ +\rangle, -\rangle\}$ Bases	72
3.5	Plot of ΔZ as a function of $ \alpha ^2$	75
4.1	Example of a Practical Multi-Level System	90
4.2	Larmor Precession on the Bloch Sphere	94
4.3	Plot of the probabilities $P_0(t)$ and $P_1(t)$ as functions of time	98
5.1	Examples of Classical Irreversible Gates	110
5.2	R_x , R_y , and R_z Gates as Rotations on the Bloch Sphere	116
5.3	Demonstration of Non-Commutative Nature of 3D Rotations	117
5.4	The Unified Rotation Gate (R_u) on the Bloch Sphere	118
5.5	Examples of Quantum Circuits	120
5.6	Quantum Random Number Generator (QRNG)	123
5.7	Fundamental BB84 QKD Protocol	126
5.8	BB84 QKD Implementation with Qubits	129
5.9	Fair Coin Gate - Setup and Rules	131

5.10	Quantum Coin Gate - Setup and Rules	131
5.11	Strategy in Which Alice Always Wins	132
5.12	Strategy Shift: Bob Beats Alice	133
5.13	Strategy Revised: Alice Wins Again	133
5.14	Illustration of the No-Cloning Theorem	134
6.1	Local Measurement Examples	146
6.2	Local Measurement in $\{ +\rangle, -\rangle\}$ Basis	148
6.3	Measurement on a Group of Qubits	159
6.4	A Partial Measurement Paradigm for Quantum Algorithms	159
6.5	Multi-Qubit Measurement in Alternate Basis	162
6.6	Illustration of Ising Interaction	165
7.1	Examples of Classically Controlled Gates	178
7.2	Parallel Gate on Unentangled State	179
7.3	Gate on Part of an Entangled State	179
7.4	Expansion of the Boolean Representation of CNOT	184
7.5	Example of Deferred Measurements	192
7.6	Deferred Measurement Principle	193
8.1	Bell State Creation using H and CNOT Gates	206
8.2	Bell State Creation from Initial State $ 00\rangle$	206
8.3	Bell State Creation Using Output Phase Shift	207
8.4	Photon Pair Creation with SPDC	207
8.5	Boolean Representation of Bell States	208
8.6	Bell Measurement	209
8.7	Bell Measurement with Post-Measurement State Restored	210
8.8	Converting $ \beta_{00}\rangle \equiv \Phi^+\rangle$ to Other Bell States	210
8.9	Converting Other Bell States to $ \beta_{00}\rangle \equiv \Phi^+\rangle$	211
8.10	Additional Conversions Related to $ \beta_{11}\rangle \equiv \Psi^-\rangle$	211
9.1	Classical Correlation versus Quantum Entanglement	219
9.2	Settings for Violating the Bell-CHSH Inequality	228
9.3	Basic Experimental Setup for Bell Test	231
9.4	Illustration for Problem 9.5	243
10.1	Bell State Generator and Analyzer	247
10.2	Basic Process of Superdense Coding	249
10.3	Basic Quantum Circuit for Superdense Coding	249
10.4	Basic Process of Quantum Teleportation	252
10.5	Fundamental Quantum Circuit for Teleportation	252
10.6	Quantum Circuit for Teleportation with Deferred Measurement	253
10.7	Basic Process of Entanglement Swapping	261
10.8	Fundamental Quantum Circuit for Entanglement Swapping	261
10.9	Basic Process of Quantum Gate Teleportation	265
10.10	Quantum Circuit for Single-Qubit Gate Teleportation	266
10.11	Gate U Applied Post-Teleportation	266
10.12	Single-Qubit Gate Teleportation with Deferred U Gate	269

10.13	Basic Process of 2-1 Gate Teleportation	271
10.14	Quantum Circuit for 2-1 Gate Teleportation	272
10.15	Gate U Applied Post-Teleportation	272
10.16	Quantum Circuit for 2-1 CNOT Teleportation	274
10.17	Basic Process of 2-2 Gate Teleportation	275
10.18	Quantum Circuit for 2-2 Gate Teleportation	275
10.19	Two-Qubit U -Gate Applied Post-Teleportation	276
10.20	Quantum Circuit for CNOT Gate Teleportation	277
10.21	Basic E91 QKD Protocol	279
11.1	Quantum Circuit for the Deutsch-Jozsa Algorithm	291
11.2	Quantum Circuit for Spectral Measurement	295
11.3	An Oracle for a Balanced Function	297
11.4	A Phase Oracle for a Balanced Function	297
11.5	A Phase Oracle Using CZ Gates	298
11.6	Max-Cut Problem Example	302
11.7	VQE Ansatz and Iteration Procedure	306
11.8	QAOA Ansatz and Iterative Procedure	308
11.9	Example of a QAOA Ansatz Building Block	310
11.10	Example of the Traveling Salesman Problem	311
11.11	Quantum Bomb Tester - Classical Scenario	316
11.12	Quantum Bomb Tester - Measuring in the $\{ +\rangle, -\rangle\}$ Basis	316
11.13	The Quantum Bomb Test Algorithm	317
11.14	Adaptive Attack on Wiesner Quantum Money	321
12.1	From NISQ to Fault Tolerance	335
12.2	Mixed State Inside the Bloch Sphere	345
12.3	Stinespring Dilation: Unitary Embedding of a CPTP Map	353
12.4	Repeated Application of an Ideal X Gate	361
12.5	Repeated Application of an Off-Calibrated X Gate	362
12.6	Effect of ZZ Cross-Talk over Time	363
12.7	Bit-Flip Error in the Bloch Sphere	365
12.8	Repeated Applications of an X Gate with Bit-Flip Noise	366
12.9	Phase-Flip Error in the Bloch Sphere	366
12.10	Depolarizing Error in the Bloch Sphere	367
12.11	Structure of Practical Error Correction Code	376
12.12	Quantum Circuit for the Three-Qubit Bit-Flip Code	377
12.13	Quantum Circuit for the Three-Qubit Phase-Flip Code	380
12.14	Quantum Circuit for the Nine-Qubit Shor Code	382
12.15	Error Syndrome Measurement and Correction for the Shor Code	383
12.16	Nine-Qubit Shor Code Simplified for Oneshot Operation	386
13.1	Venn Diagram for Probability Relationship	405
13.2	Quantum Simulation of 1D Random Walk	415
13.3	Quantum 1D Walk	416
13.4	Venn Diagram for Shannon Entropy Relationships	422
13.5	Venn Diagram for von Neumann Entropy Relationships	426
13.6	Schematic of the Quantum Information Transmission Process	433

List of Tables

1.1 Key Photon Polarization States	8
1.2 Summary of Photon Polarization States and Observables	22
2.1 Special Spin States	47
2.2 Spin State and Polarization State Correspondence	48
3.1 Examples of Qubit Implementations	58
5.1 Common Single-Qubit Quantum Gates	112
5.2 Transformations by Common Single-Qubit Quantum Gates	114
7.1 Common Multi-Qubit Quantum Gates	172
7.2 Boolean Representation of Common Qubit Gates	183
7.3 Gate Sequence Equivalence Involving X , Z and CNOT	189
7.4 Gate Sequence Equivalence Involving H	190
7.5 Expressing Common Gates Using Toffoli Gate	195
10.1 Summary of Key Applications of Quantum Entanglement	246
10.2 Correction Gates for General Qubit States	247
10.3 Conversion Gates for Bell States	248
10.4 Boolean Representation of Selected Qubit Gates	254
10.5 Gate Sequence Equivalence for Two-Qubit Gate Teleportation	276
10.6 Gate Sequence Equivalence for CNOT Teleportation	277
11.1 Hadamard Transform H_4 and Period Functions	293
11.2 Relationships Between Terms Related to Quantum Optimization	301
12.1 Error Syndrome and Correction for the Three-qubit Bit-flip Code	379
12.2 Error Syndrome and Correction for the Three-Qubit Phase-Flip Code	381
12.3 Error Syndrome for the Nine-Qubit Shor Code	385
13.1 Evolution of Coin & Walker States in a 1D Quantum Walk	417

Index

A

- adiabatic quantum computing 92, 312
anyons 42, 157

B

- basis equivalence 21
basis states 11
basis-dependent measurement 72
Bell analyzer 209
Bell basis 203
Bell inequalities 218, 223
 Bell test 230
 derivation 235
 experimental verification 230
 measurement bases 235
Bell measurement 208
Bell states 201
 conversion 209
 creation 205
 definitions 202
 generalization 212
 maximum entanglement 203
 orthonormal basis 203, 211
 summary 246
Bell test
 early experiments 231
 experimental setup 230
 findings and implications 232
 loophole-free experiments 232
Bell-CHSH inequality 223
 classical cases 225
 experimental verification 230
 quantum cases 227
Bloch sphere 48, 60
 bit-flip error 364
 depolarizing error 367
 general qubit states 60
 key states 49
 mixed states 344
 Pauli X, Y, Z gates 49

- phase-flip error 366
rotation gates 116
Boolean representation 182
 caveat 184
 common gates 183
 interpretation 184
 properties 182

C

- change of basis 62
 eigenbasis 67
 general qudits 65
 Hadamard basis 63
classical correlation 218
classical random walk 414
classically controlled-U gate 177
CNOT gate 173
 flipped version 174
 negative version 177
CNOT gate implementation 165
code distance 386
coherent errors 361
 cross-talk error 362
 gate calibration error 361
common single qubit gates 111
compatible observables 33
composite system postulate 152
computation in place 121
conditional entropy 422, 425
conditional probability 404
continuous-time quantum walk 417
CPTP maps 350

D

- data processing inequality (DPI) 429
decoherence by measurement 348
deferred measurement principle 192
density operator 338
 measurements 348
 partial trace 354
POVM measurements 350

density operator:Bloch sphere 344
density operator:time evolution 346
Deutsch-Jozsa algorithm 290
Dirac notation 21

E

E91 QKD protocol
comparison to BB84 280
eavesdropping check 280
key sifting 280
procedure 278
empirical expectation value 74
energy eigenstates 88
energy levels 88
entanglement 218, 219
applications 246
misconceptions 234
vs. classical correlation 218
entanglement swapping
analysis 261
Bell measurement 262
classical communication 263
initial entanglement 262
procedure 260
state reconstruction 263
EPR paradox 221
locality 222
realism 222
equivalent gate sequence 186
analysis approaches 187
CNOT gate sequences 189
H gate sequences 190
single-qubit gates 121
error correction codes 376
error syndrome measurement 383
excited states 89
expected (expectation) value 28

F

fault-tolerant quantum computing 336
fermions and bosons 42

G

gate-based quantum computing 91
general qubit state 59
general qudit state 61
generalized DJ algorithm 293

GHZ state 155, 359
GHZ states 212
graph coloring problem 330
ground state 89

H

Hadamard basis 63
Hadamard gate 114
hidden variable theory 222
Holevo (HSW) theorem 431
Holevo's bound 431
holonomic quantum computation 390

I

incoherent errors 363
amplitude damping error 368
bit-flip error 364
dephasing error 367
depolarizing error 367
Pauli channel error 368
phase-flip error 366
two-qubit bit-flip error 368
two-qubit dephasing error 368
two-qubit Pauli channel 368
informationally secure vs. computationally secure 323
Ising model 164, 303

J

joint entropy 421, 425
joint measurements 149
joint probability 404

K

Kraus decomposition 351

L

Larmor precessing 93
lattice-based cryptography 326
logical gates 382

M

Mach-Zehnder interferometer 30

magnetic moment of spin	43
marginal probability	404
max-cut problem	300
measurement basis	73
measurement errors	369
measurement framework	69
measurement probability	23
measurement state collapse	25
mixed state	339
monogamy of entanglement	359
multi-qubit gates	170
CNOT (Controlled NOT)	173
Controlled-U	175
CZ (Controlled Z)	178
interaction gates	178
ZZ gate	178
multi-qubit systems	138, 151
basis states	152
entangled states	155
Hamiltonian	162
measurements	157
product states	154
time evolution	162
mutual information	423, 426

N

nine-qubit Shor code	381
NISQ era	335
NISQ/hybrid algorithms	298
no-cloning theorem	110, 134, 359
no-communication theorem	234, 359
Nobel prize in physics, 2022	233
number partitioning problem	329

O

operator diagonalization	67
operator-sum representation	351
orthogonal qubit state	59

P

parallel gates	178
parity measurements	386
partial measurement	158
partial product notation	158
partial trace	354
Pauli gates	113
performance metrics	369

algorithmic qubits (AQ)	374
application footprint	375
circuit depth	373
CLOPS	374
coherence time (T_2)	370
coherence to gate time ratio	371
composite metrics	373
connectivity	371
cross-talk	372
error rate	373
fidelity	372
gate time	371
QECC decoding complexity	375
quantum volume (QV)	373
qubit lifetime	371
relaxation time (T_1)	370
scalability metrics	375
subthreshold scaling	375
the Λ factor	375
phase gates	115
phase kickback	297
phase oracle	296
photon circular polarization	14
photon cross-polarizer experiment	27
photon diagonal polarization	12
photon double-slit experiment	32
photon elliptical polarization	15
photon polarization	6
photon polarization observables	19
photon rectilinear polarization	10
photons	6
post-measurement ensemble	348
post-quantum cryptography	326
POVM measurements	71, 350
principle of entropy increase	429
projective measurement	71
pure state	338

Q

QAOA algorithm	307
QM postulates	4
composite systems	152
quantum measurements	23
quantum observables	16
quantum states	16
time evolution	83
quantum algorithm categories	288
quantum annealing	92, 312
quantum bomb test algorithm	315
quantum channel	351

- quantum channel capacity 431
quantum circuit diagram 120
quantum communication
 BB84 QKD protocol 124
 channel capacity 431
 cryptography 326
 E91 QKD protocol 277
 entanglement swapping 260
 entropy 417
 entropy increase principle 429
 quantum gate teleportation 264
 quantum teleportation 251
 superdense coding 248
quantum error correction 333
quantum gate teleportation
 Bell measurement 267
 classical communication 268
 Clifford group 269
 CNOT gates 273, 275
 from standard teleportation 265
 initial state 267
 Pauli gates 269, 270
 single-qubit gate 265
 state reconstruction 268
 two-qubit gates 271, 273
quantum information 396
quantum interference 30
quantum measurement 68
quantum measurement postulate 23
quantum money 318
 adaptive attack 321
 challenge-and-response prot. 323
 learning attack 320
 private key protocol 319
 public key implementation 324
 public key protocol 323
 replication attack 320
 Wiesner quantum money 319
quantum observable postulate 16
quantum probability 399, 401
 distinctive features 399
 entangled states 410
 expected values 403
 joint distribution 406
 marginal probability 407
 multi-variable 406
 non-commuting observables 407
 observables 401
 probability distribution 402
 random variables 403
quantum random number generator 122
quantum state postulate 16
quantum state tomography 77
quantum stochastic process 412
quantum teleportation
 basis state approach 252
 Bell measurement 257
 Bell measurement approach 256
 classical communication 258
 deferred measurement 253
 establish entanglement 253, 255
 logic operation approach 254
 procedure 251
 state reconstruction 253, 256, 258
quantum thermodynamics 430
quantum utility era 335
quantum walk 415
qudit framework 56
- R**
- Rabi oscillations 96
readout errors 369
reduced density operator 354
rotation gates 116
- S**
- sampling errors 74
Schrödinger equation 84, 162
Shannon entropy 418
SPDC photon pair 207
spin 42
spin measurement 49
spin precessing 93
spin-1/2 particles 42
stabilizers 386
state preparation errors 369
stationary states 88
statistical average 28
Stern-Gerlach experiment 50
 cascaded 52
Stinespring dilation 353
subadditivity 434
superdense coding
 decoding 250
 encoding 249
superluminal communication 234
superposition 11

T

the circuit model	119
three-qubit bit-flip code	377
three-qubit phase-flip code.....	380
time evolution of qubit systems	82
operator	84
postulate.....	83
Toffoli (CCNOT) gate	195
topological quantum computation .	42,
157, 390	
trace preserving operations	350
traveling salesman problem	310
Trotterization.....	87
two-path interference	30
two-qubit rotation gates	178
two-qubit systems	139
entangled states.....	142
general states	142
joint measurements	149
local measurements	143
measurements.....	143
parity measurements.....	149
product states	142

U

uncertainty inequality	34
uncertainty principle.....	32
unified rotation gate.....	118
unitary evolution properties	84
universal gate set	179
Clifford group + T	181
Toffoli + H	181
universal quantum computing.....	91

V

von Neumann entropy	420
strong subadditivity.....	435
von Neumann equation	347
VQE algorithm	304

W

W state	359
---------------	-----

Journey Forward

A Journey Well Traveled

Congratulations to all of you who have navigated the complexities of this book, with a special nod to those who delved into the exercises. You've successfully laid a solid foundation in quantum computing and quantum information science, equipping yourselves for further endeavors in quantum exploration.

The Quantum Horizon

As with the dawn of classical computing, we stand at a significant juncture in the evolution of quantum computing. However, the path ahead is not without its challenges. Achieving scalability, mitigating noise, and ensuring fault tolerance remain the holy grails of the field. While we have yet to fully realize these goals, the quantum community remains optimistic, knowing that transformative breakthroughs may be just around the corner.

Exploring quantum algorithms opens doors to myriad applications in optimization, machine learning, and quantum simulations. However, their full potential will be unlocked only alongside advancements in quantum hardware, with technologies like superconducting qubits, neutral atoms, and trapped ions leading the charge.

Confronted with the challenges of decoherence and noise, research into quantum error correction becomes indispensable, ensuring the transition from theoretical ideas to practical, fault-tolerant quantum operations. Beyond the immediate realm of computation, disciplines such as quantum information theory, sensing, and post-quantum cryptography enrich the broader quantum narrative, providing layered insights into this ever-evolving domain.

To remain relevant in this dynamic field, staying updated is crucial. From groundbreaking work in quantum cryptography to the emergence of innovative algorithms and their applications, the story of quantum science is continually unfolding. Armed with the foundational knowledge from this book, you are well-positioned to both witness and contribute to the future chapters of this quantum tale.

Looking Forward

As you progress, consider:

- Delving deeper into quantum algorithms, which offer solutions across varied application domains.
- Keeping pace with the latest developments in quantum computing, including hardware innovations, error correction methods, advancements in cryptography, and emerging algorithms.
- Exploring the broader canvas of quantum information science, where disciplines like quantum sensing and post-quantum cryptography add additional depth.

In conclusion, while this book provides a sturdy foundation, the journey is ongoing. With relentless curiosity and continuous learning, traverse the nuances of the quantum world. Wishing you a quantum voyage that is as enlightening as the principles upon which it is built.

If you've found value in this quantum computing journey,
please consider leaving a rating and review at
<https://amazon.com/gp/product-review/B0CY3JJPNB>.

Your feedback supports our work and helps others discover
this fascinating field. Thank you!

многоэлектронные атомы

в нейтральном атоме число протонов = числу электронов в ядре

атом = положительно заряженное ядро + группа локализованных вблизи ядра электронов

$$\hat{H} = \hat{T} + \hat{V} + \hat{W} + \hat{Q}$$

$$\hat{T} = \sum \frac{\hat{p}_k^2}{2\mu} = - \sum \frac{\hbar^2}{2\mu} \nabla_k^2$$

\hat{V} - эенргия притяжения всех электронов

$$\hat{V} = - \sum \frac{Ze^2}{r_k}$$

- электростатическое напряжение электронов в ядре

$$\hat{W} = \frac{1}{2} \sum_{k,l=1, k \neq l} \frac{e^2}{|r_k - r_l|}$$

- электростатическое отталкивание

Q - ?

все типы взаимодействия которые могут возникать из-за спина = Q

магнитное поле одного электрона действует на спин другого электрона, спиновый магнитный момент создает магнитное поле, взаимодействует со спином первого

$$\hat{H}\Psi_{n\lambda} = E_{n\lambda}\Psi_{n\lambda}, \quad n = 1, 2, \dots \quad \alpha^2 = \frac{e^2}{\hbar c}$$

$\|Q\|$ меньше порядка α^2 других слагаемых

появится дискретный спектр энергии при решении уравнения, уровень не определяет энергию, просто классификация уровней, есть какие-то иные физические переменные, обозначим его за λ

Модель невзаимодействующих электронов

$$\hat{W} = 0, \hat{Q} = 0$$

нет ни одного атома, для которого существует точное аналитическое решение уравнения Шредингера. Существует простая модель многоэлектронного атома, основанная на предположении, что вклад слагаемых \hat{W}, \hat{Q} в оператор \hat{H} в определенном смысле не сравним с \hat{T}, \hat{V}

$$\begin{aligned} \hat{H}^0 &= \hat{T} + \hat{V}; \quad \hat{H}^0 \Phi_{n\lambda} = E_{n\lambda}^0 \Phi_{n\lambda} \\ \hat{H}^0 &= \sum \left[-\frac{\hbar^2}{2\mu} \nabla_k^2 - \frac{Ze^2}{r_k} \right] = \sum \hat{h}_k^0(r_k) \\ \epsilon_n &= -\frac{Z^2}{2n^2} \frac{e^2}{a_0}, \quad n = 1, 2, 3, \dots \end{aligned}$$

, $a_0 = \frac{\hbar}{2}$ - боровский радиус

$$E_{n\lambda} = \sum_k \epsilon_{n,k}$$

модель невзаимодействующих электронов при описания состояний многоэлектронного атома предполагает, что эти состояния включают пречисление одночастичных квантовых чисел электронов системы.

Оператор перестановки частиц

$P_{ij}F(x_1, x_i, \dots, x_j, \dots, x_N) \rightarrow P_{i,j}F(x_1, x_j, \dots, x_i, \dots, x_N)$ меняет местами координаты с индексами i и j
 $P\Psi = \lambda\Psi \rightarrow P^2\Psi = P\lambda\Psi$

$$P^2\Psi = \Psi = \lambda P\Psi = \lambda^2\Psi, \quad \lambda = \pm 1$$

оператор перестановки коммутирует с $\hat{V}, \hat{W}, \hat{Q}$

Тождественные и не тождественные частицы

Тождественные - все свойства одинаковые (заряд, спин, масса), электроны - тождественные частицы

Оператор перестановки пространственно-спиновых переменных коммутирует с оператором Гамильтона

Квантовый интеграл движения, входит в набор переменных задающих состояние

наличие оператора перестановки предположило, что для электронов входящих в состав атома реализуются только антисимметричные состояния

принцип Паули - две частицы не могут иметь одни и те же квантовые числа (эквивалентно выше)

О волновой функции многоэлектронного атома одночастичные собственные функции оператора $h_k^0(\vec{r}_k)$ используется для нахождения явного вида многочастичной волновой функции.

Набор одночастичных квантовых чисел включает в себя главное квантовое число n орбитальное квантовое число l и проекцию спинового момента m . предполагают задание только n и m включение в описание операторов, опущенных в модели невзаимодействующих электронов, приводит к появлению новых многочастичных квантовых чисел. Так учет электростатического взаимодействия приводит к введению новой величине - суммарному моменту всех электронов $\hat{L} = \sum_k \hat{l}_k$. оператор \hat{L}^2 коммутирует с атомным гамильтонианом, в котором учтено межэлектронное взаимодействие. Оператор Q позволяет ввести суммарный спин $\hat{S} = \sum_k \hat{s}_k$ как дополнительное описание состояний. Суммарный спин коммутирует с атомным гамильтонианом в приближении невзаимодействующих электронов. Оба же оператора приводят к появлению оператора полного момента атома $\hat{J} = \hat{L} + \hat{S}$, квадрат которого коммутирует с полным оператором гамильтона для атома ...

Используя отмеченные условности, приведем несколько примеров описания состояний многоэлектронных атомов, не решая уравнения Шредингера. Следовательно, в модели невзаимодействия электронов состояние задается конфигурацией $n_1 l_1, n_2 l_2$. При этом состояние с минимальной энергией задается конфигурацией $1s1s$. С расширенными приближенными квантовыми числами в этом состоянии суммарный орбитальный момент $L = 0$, суммарный спин $S = 0$ и полный момент атома $J = 0$ также равен нулю

Термы

$$2S+1 L_j$$

, где $2S + 1$ = мультиплетность.

Состояние термов со значением орбитального момента $L = 0, 1, 2, 3, 4, 5, \dots$ обозначается буквами S, P, D, F, G, H, ... соответственно. Термы с мультиплетностью = 1 называются синглентными

Так, возвращаясь к атому гелия, можно сказать, что его основному состоянию соответствует терм 1S_0

Следующим за атомом гелия по числу электронов является атом лития. Конфигурация основного состояния $1s^2 2s$ и терм $^2S_{1/2}$. Конфигурация возбуждения $1s^2 nl$ определяется набором дублетных термов $n^2 S_{1/2}, n^2 P_{1/2}, n^2 P_{3/2}$ и т.д. Аналогичный набор атомных термов возникает у всех щелочных атомов (Li, Na, K, Rb, Cs, Fr) с одним электроном вне замкнутой оболочки. Замкнутые оболочки имеют значение суммарных

Группа атомов с двумя внешними электронами Be, Mg, Ca, Sr, Ba, Ra образует группу щелочеземельных металлов, система атомных термов которых схожа с набором термов атома Не

"Охлаждение" атомов

охлаждающие лазеры с нескольки стороны для удержания атомов на месте

Задача удержания атомов в конкретном положении

Молекула

Энергия \rightarrow электрон - свое ядро + электрон чужое ядро + электрон свой чужой

Квантовая теория представляет молекулу как систему, состоящую из нескольких атомов. Таким образом, молекула состоит из ядер атомов и электронов, локализованных в области близости ядер. У молекул нет объекта типа атома водорода, знание которого позволяет построить разумную модель на простом базисе. Оператор гамильтона слишком сложен, чтобы отыскать аналитическое решение уравнения Шредингера. Квантовая теория предполагает, что у этих систем есть дискретные квантовые состояния и имеется вероятность перехода между этими состояниями.

Набор электронных состояний молекулы определяется свойствами электронной оболочки и симметрией молекулы

ЯМР(ядерный магнитный резонанс) - явление резонансного поглощения радиочастотной электромагнитной энергии веществом с ненулевыми значениями магнитных моментов ядер, находящимся во внешнем постоянном магнитном поле

Например, ненулевыми ядерными магнитными моментами обладают ядра 1H , 1C , 1F , 3P . ЯМР наблюдается в однородном постоянном магнитном поле, на которое накладывается слабое переменное радиочастотное поле, перпендикулярное постоянному. Для веществ, у которых спин равен $1/2$, в постоянном поле возможны две ориентации магнитного момента ядра. При условии, что частота радиочастотного поля, умноженная на постоянную Планка, равно разности этих уровней энергии, наблюдается резонансное поглощение энергии радиочастотного поля, названное ЯМР.

Агрегатные состояния вещества

Газ

Жидкость

твердое тело

Плазма

если электронейтральные объекты возникает плазменное состояние

Квантовые информационные системы могут быть реализованы на любом агрегатном состоянии
Прохождение излучения через вещество

$$I = I_0 e^{-\mu x}$$

$$I = I_0 \exp - \int_L \mu x$$

Лазеры. Мазеры

Инверсивная заселённость - это состояние квантовой системы, при котором населённость верхнего энергетического уровня больше населённости нижнего энергетического уровня

Нормальная заселённость - большинство атомов находятся в основном состоянии

$$\mu = \alpha(n_1 - n_2)$$

, n_1, n_2 - число атомов в основном и возбужденном состояниях

При работе лазера часто используется система трёх энергетических уровней атома, второе из которых - метастабильное со временем жизни атома в нем до 10^{-3} с

2-й уровень является метастабильным по отношению к 4
во 2 легко попасть, из 4 легко удалиться => во 2 накапливается => инверсивная заселённость

реальные устройства используют явление сверхпроводимости

сверхпроводимость - свойство некоторых материалов проводить электрический ток без диссипации с одновременным выталкиванием магнитного поля, - явлением известным как эффект Мейснера, заключающийся в полном или частичном вытеснении магнитного поля из объема сверхпроводника. Существование этого эффекта показывает, что сверхпроводимость не может быть описана как идеальная проводимость в классическом понимании

Эффект Джозефсона в сверхпроводниках - явление протекания сверхпроводящего тока через тонкий слой диэлектрика, разделяющий два проводника. Такой ток называют джозефсоновским током, а также соединением сверхпроводников - джозефсоновским контактом

Квантовые точки

Квантовая точка - фрагмент проводника или проводника, носители заряда (электрона или дырки) которого ограничены в пространстве по всем трём измерениям. Размер квантовой точки должен быть настолько мал, чтобы квантовые эффекты были существенными

Ионная ловушка

В ловушке Пола осевое движение ограничено статическим полем, а радиальное колебательным.

Квантовая модель вычислений

0 - низкий потенциал, 1 - высокий потенциал

Логические элементы (операторы = вентили = гейты) проеобразования одного кубита

Скрепленными магнитными полями можно управлять состоянием спина электрона.

В общем случае описание преобразования одного кубита можно формализовать заданием графических обозначений для логических элементов и их последовательного действия (квантовая цепь)

Физически однокубитовыми операциями являются логические элементы \hat{U} , на вход которых подается кубит в некотором нач. состоянии $|\psi\rangle$, а на выходе данного элемента появляется проеобразованный кубит.

Квантовый регистр

Совокупность заданного числа кубит
число состояний в регистре - 2^N

2-х кубитовый регистр

$$|q\rangle_2 = |q_1 q_2\rangle = |q_1\rangle |q_2\rangle = (\alpha |0\rangle + \beta |1\rangle) \otimes (\alpha' |0\rangle + \beta' |1\rangle) = \\ = c1(|0\rangle \otimes |0\rangle) + c2(|0\rangle \otimes |1\rangle) + c3(|1\rangle \otimes |0\rangle) + c4(|1\rangle \otimes |1\rangle)$$

$$|0\rangle_2 = |00\rangle = |0\rangle \otimes |0\rangle$$

$$|1\rangle_2 = |01\rangle = |0\rangle \otimes |1\rangle$$

$$|2\rangle_2 = |10\rangle = |1\rangle \otimes |0\rangle$$

$$|3\rangle_2 = |11\rangle = |1\rangle \otimes |1\rangle$$

$$|q\rangle_2 = \sum_{k=0} c_k |k\rangle_2, \quad \sum_{k=0} |c_k|^2 = 1$$

$$|0\rangle_2 = |0_a 0_b\rangle = |0_a\rangle \otimes |0_b\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$|1\rangle_2 = |0_a 1_b\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$|q\rangle_n = \sum_{k=0}^{2^n-1} c_k |k\rangle_n, \quad \sum_{k=0}^{2^n-1} |c_k|^2 = 1$$

$$|0\rangle_n = |00\dots 00\rangle = |0\rangle \otimes |0\rangle \otimes \dots \otimes |0\rangle \otimes |0\rangle$$

$$|2^n-1\rangle_n = |11\dots 11\rangle = |1\rangle \otimes |1\rangle \otimes \dots \otimes |1\rangle \otimes |1\rangle$$

Измерение регистра

Если система находилась в состоянии суперпозиции $\Psi = \alpha |0\rangle + \beta |1\rangle$, то в результате ее мы получим вектор $|0\rangle$ с вероятностью $|\alpha|^2$, а вектор $|1\rangle$, с вероятностью $|\beta|^2$

Однокубитовые операторы

$$\text{NOT} : |\psi\rangle = \text{NOT} : (a|0\rangle + b|1\rangle) = a|1\rangle + b|0\rangle$$

$$\psi = (a/b) -> a|0\rangle + b|1\rangle; \quad |a|^2 + |b|^2 = 1$$

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, X \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} b \\ a \end{pmatrix}$$

можно построить неограниченное число однокубитовых операторов. Однако в силу полноты системы матриц Паули и Единичной матрицы I любая 2×2 матрица м.б. разложена по этой полной системе матриц.

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}; \quad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}, \quad P(c, d) = \begin{pmatrix} e^{ic} & 0 \\ 0 & e^{id} \end{pmatrix}$$

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$Z|\psi\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = a|0\rangle - b|1\rangle$$

Квантовый гейт Адамара

$$H = \frac{1}{\sqrt{2}}(\sigma_x + \sigma_z) = \frac{1}{\sqrt{2}}(X + Z)$$

$$H : |0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}, \quad H : |1\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$$

Однако квадрат гейта Адамара $H^2 = H \cdot H$ не приводит к квантовому NOT-гейту, т.к. $H^2 = I$ и поэтому H не является квадратным корнем от гейта NOT = X. Двукратное применение гейта H возвращает систему в исходное состояние. В этом смысле гейт Адамара является обратимым однокубитовым гейтом

$$\begin{aligned} \sqrt{NOT} &= \sqrt{X} = \frac{1}{2} \begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix} \\ \sqrt{NOT}\sqrt{NOT}|0\rangle &= |1\rangle, \quad \sqrt{NOT}\sqrt{NOT}|1\rangle = |0\rangle \end{aligned}$$

Операторы поворота Rx, Ry, Rz

Гейт rx на сфере блоха соответствует вращению кубита вокруг оси X на заданный угол. В матричном виде можно записать как

$$Rx(\theta) = e^{-i\frac{\theta}{2}X} = \begin{pmatrix} \cos(\frac{\theta}{2}) & -i\sin(\frac{\theta}{2}) \\ -i\sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{pmatrix}$$

Операторы U1 U2 U3

$$U1(\lambda) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\lambda} \end{pmatrix}$$

$$U1(\pi) = Z$$

$$U1(\pi/2) = S$$

$$U1(\pi/4) = T$$

$$\begin{aligned} U2(\phi, \lambda) &= Rz(\phi + \frac{\pi}{2})Ry(\frac{\pi}{2})Rz(\lambda - \frac{\pi}{2}) = \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -e^{i\lambda} \\ e^{i\phi} & e^{i(\phi+\lambda)} \end{pmatrix} \\ U3(\theta, \phi, \lambda) &= Rz(\phi - \pi/2)Rx(\frac{\pi}{2})Rx(\pi/2)Rz(\lambda - \pi/2) \\ U3 &= \begin{pmatrix} \cos(\frac{\theta}{2}) & -e^{i\lambda}\sin(\frac{\theta}{2}) \\ e^{i\phi}\sin(\frac{\theta}{2}) & e^{i(\phi+\lambda)}\cos(\frac{\theta}{2}) \end{pmatrix} \end{aligned}$$

$$CNOT = \begin{array}{c} |i\rangle \xrightarrow{\text{---}} \bullet \xrightarrow{\text{---}} |i\rangle \\ |j\rangle \xrightarrow{\text{---}} \boxed{\text{NOT}} \xrightarrow{\text{---}} |i\oplus j\rangle \end{array} \stackrel{=}{\sim} \begin{array}{c} |i\rangle \xrightarrow{\text{---}} \bullet \xrightarrow{\text{---}} |i\rangle \\ |j\rangle \xrightarrow{\text{---}} \bigoplus \xrightarrow{\text{---}} |i\oplus j\rangle \end{array}$$

Многокубитовые квантовые гейты

$$CNOT : |k\rangle_2 = CNOT : |i, j\rangle = |i, i\oplus j\rangle$$

$$CNOT : \sum_{k=0}^3 c_k |k\rangle_2$$

гейт Уолша-Адамара

$$\hat{W}^{(n)} = H_1 \otimes H_2 \otimes \dots \otimes H_n$$

$$\hat{W}^{(n)} : |0\rangle_n = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} |k\rangle_n$$

$$U : \hat{W}^{(n)} : |0\rangle_{n+1} -> \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} |k\rangle_n \otimes |f(k)\rangle$$

$$\hat{W}^{(n)} |i\rangle = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} (-1)^{i \cdot k} |k\rangle_n$$

$$i \cdot k = \sum_{m=0}^{2^n-1} (i_m \wedge k_m) = i_0 \cdot k_0 \oplus i_1 \cdot k_1 \oplus \dots$$

Преобразование регистров

$$|Q\rangle_n = \sum_{k=0}^{2^n-1} c_k |k\rangle_n = \sum_{k_1, k_2, \dots, k_n \in \{0, 1\}} C_{k_1, k_2, \dots, k_n} |k_1, k_2, \dots, k_n\rangle$$

В таких обозначениях связь более компактных десятичных обозначений с бинарными обозначениями означает, что $c_0 = C_{0,0,\dots,0}$, $c_1 = C_{0,0,\dots,1}$ и т.д. для амплитуд базисных состояний и аналогично для самих базисных состояний $|0\rangle_n = |0, 0, \dots, 0\rangle$, $|1\rangle_n = |0, 0, \dots, 1\rangle$ и т.д.

$$\sigma_x^{(j)} : |Q\rangle_n = \sum_{k_1, k_2, \dots, k_j, \dots, k_n \in \{0, 1\}} C_{k_1, \dots, k_j, \dots, k_n} |k_1, \dots, k_j, \dots, k_n\rangle$$

$$\text{двукубитовый регистр } \sigma_x^{(2)} : |Q\rangle_2 = \sigma_x^{(2)} : \sum_{k=0}^3 c_k |\vec{k}\rangle_2 = \sigma_x^{(2)} : \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_0 \\ c_3 \\ c_2 \end{pmatrix}$$

Фурье-преобразование

Классическое дискретное преобразование Фурье обычно определяется как преобразование набора $(x_0, x_1, \dots, x_{N-1})$ N -комплексных чисел в набор комплексных чисел $(y_0, y_1, \dots, y_{N-1})$

матричная форма

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \dots \\ y_{N-1} \end{pmatrix} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & q & q^2 & \dots & q^{N-1} \\ 1 & q^2 & q^4 & \dots & q^{2(N-1)} \\ \dots & & & & \\ 1 & q^{N-1} & q^{2(N-1)} & \dots & q^{(N-1)^2} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \dots \\ x_{N-1} \end{pmatrix} \text{ где } q = \exp(i2\pi/N)$$

Квантовое Фурье-Преобразование

$$|\mathbf{j}\rangle_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \exp(i \frac{2\pi}{N} jk) |k\rangle_n \quad |k\rangle_n = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} q^{j*k} |\mathbf{j}\rangle_n$$

Квантовые алгоритмы

Задача Дойча

пусть, например, имеется четыре устройства, выполняющих четыре разных преобразования, которые описываются четыремя бинарными функциями $f_i(x), i \in 1, 2, 3, 4$ от двоичной переменной $x = 0, 1$. При этом функции f_1f_2 - постоянные и принимают значения $f_1(x) = 0, f_2(x) = 1$, а две другие функции f_3f_4 определены следующим образом: $f_3(x) = x, f_4(x) = NOT(x)$. При этом функции f_3f_4 называются сбалансированными. Требуется определить, к какой группе относятся функция преобразования какого-то из этих четырех устройств, если исходно это неизвестно.

Задача Дойча-Джозса

Задача в данном случае формулируется подобно задаче Дойча и имеет целью определить, является ли бинарная функция $f(x) f \in 0, 1$, где x - многомерный вектор, постоянной или сбалансированной. Под постоянством функции понимается выполнения равенств $f(x) = 0$ или $f(x) = 1$ при произвольных x , а сбалансированность означает, что функция принимает значение 0 для одной половины данных x и значение 1 для другой половины данных . Квантовый алгоритм Дойча-Джозса позволяет решить эту задачу за n операций, тогда как классический алгоритм имеет решение только за 2^n операций.

Алгоритм может быть продемонстрирован на примере следующей задачи. Абонент А, находясь в некотором месте, выбирает произвольное число x из множества $0 \div 2^n - 1$ двоичных чисел и посыпает его по почте абоненту Б, находящемуся в другом месте. Абонент Б вычисляет некоторую функцию от полученного значения $f(x)$ и сообщает абоненту А результат вычислений, который может быть 0 или 1, почтовым отправлением в адрес абонента А. При этом абонент Б имеет возможность использовать функцию $f(x)$, одну из двух типов. Задача абонента А состоит в том, чтобы по полученной от абонента Б информации достоверно определить, затратив при этом минимум корреспонденции, какой тип функции выбран абонентом Б для вычислений.

Алгоритм Бернштейна-Вазириани

Задача состоит в определение бинарной строки s по результату вычисления функции следующего вида (где s^*x сумма побитового произведения по модулю 2, где $s = (s_0, s_1, \dots, s_n)x = (x_0, x_1, \dots, x_n)$) $f_s(x) = s \cdot x \bmod 2$ Классическое решение состоит в последовательном вычислении побитового произведения x на $(100\dots0)$ - определение s_0 , на $(010\dots0)$ - определение s_1 и т.д.

Алгоритм Саймона

Алгоритм Саймона сформулирован для решения задачи определения периода функции с использованием принципа интерференции квантовых состояний. Пусть есть двоичная функция $f : (0, 1)^n \rightarrow (0, 1)^n$, определенной на множестве двоичных векторов x . Функция является периодической в том смысле, что $f(x) = f(x \oplus r)$, где r - двоичный период. Путем вычисления функции найти период.

В рамках классических вычислений для решения такой задачи потребуется выполнить экспоненциальное число вычислений такой функции для сравнения значений во всех её точках. Квантовый алгоритм Саймона позволяет найти период всего за $O(n)$ вычислительных процедур. Для получения такого результата определим унитарный оператор вычисления функции f стандартным образом: $\hat{U} : |x\rangle_n \otimes |y\rangle_n \rightarrow |x\rangle_n \otimes |y \oplus f(x)\rangle_n$

Возврат фазы

В соответствии с определением гейта CNOT справедливы следующие равенства:

$$CNOT : |x\rangle \otimes \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right) = |x\rangle \otimes (-1)^x \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right) = (-1)^x |x\rangle \otimes \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right)$$

Оценка собственного значения унитарного оператора

Пусть имеется унитарный оператор U , обственный вектор которого есть $|u\rangle_m$. Собственное число любого унитарного опретора по модулю равно единице, поэтому оно может быть представлено в виде $\exp(i2\pi\phi_u)|u\rangle_m$. Таким образом, задача определения собственного числа унитарного оператора сводится к определению "фазового множителя" ϕ_n , что аналогично задаче оценки фазы.

Алгоритм Шора

Известная задача факторизации целых чисел состоит в определении простых множителей p и q для заданного целого числа $N = p^*q$. Классическое решение такой задачи опирается на алгоритм нахождения порядка числа x в арифметике по модулю. Порядком числа x по модулю N называется наименьшее натуральное число r , для которого выполняется сравнение

$$x^r \bmod N = 1$$

При этом число x удовлетворяет условию $x < N$ и, кроме того, числа x , N взаимно просты. Взаимная простота двух чисел означает, что наибольший общий делитель равен 1.

$$N = c_1 \cdot c_2, c_1 = x^{r/2} + 1, c_2 = x^{r/2} - 1$$

Алгоритм Гровера

Пусть есть неупорядоченная база данных, содержащая N элементов: x_1, x_2, \dots, x_N . Этими элементами могут быть любые сущности. Если требуется найти в этой базе конкретный элемент, который обозначим через z , то процедура отыскания этого элемента будет состоять в случайному переборе элементов базы x_i и сравнении их с образцом для поиска.

Очевидно, что если размер базы N , то может потребоваться $O(N)$ испытаний для отыскания конкретного элемента. Обозначим амплитуду вероятности извлечения x_i -го элемента через a_i и допустим, что вероятность $p_i = |a_i|^2$.

Амплитудное усиление

Выполним циклически три операции, определённые следующим тремя шагами, и проследим за изменением величины амплитуд извлечения элементов из базы:

1-й шаг: вычисление среднего значения амплитуды $\langle a \rangle$ для всех элементов базы данных:

$$\langle a \rangle = \frac{1}{N} \sum_{i=1}^N a_i$$

2-й шаг : преобразование амплитуд в соответствии с выражением: $a_i = 2 \langle a \rangle - a_i$

3-й шаг: изменение знака амплитуды искомого элемента в базе данных на противоположный: $a_p \rightarrow -a_p$