Some Mathematical Machinery
Contents

Dirac’s Ingenious Idea ........................................... 5
The Pauli Operators ............................................. 8
The Born Rule ..................................................... 14
Noncommutativity and Uncertainty ......................... 17
Entanglement ...................................................... 18
Quantum Dynamics ............................................. 21
Relative Phase .................................................... 27
Informationally Complete Observables ..................... 28
The Biorthogonal Decomposition Theorem ................. 30
States and Operators as Matrices ......................... 35
Mixed States ..................................................... 38
I show how quantum states of single systems and complex multipartite systems can be represented in a now standard notation invented by Dirac, how observables can be represented by noncommuting operators, and how multipartite systems can be in entangled states.

The discussion is not as daunting as you might expect, but if you are encountering the joys of the Dirac notation for the first time, you might need a pen and paper to play with the formalism as you read it through. A formal expression is opaque if you are not already familiar with the notation and you read it symbol by symbol instead of seeing it as a concise way of saying something that would be clumsy to say in words. So the temptation is to read through the text and skip the equations and other bits of displayed formalism. I’ll talk through the formal constructions in some detail rather than leaving it to you to figure them out. Think of it as getting a grip on the software running the universe.

**Dirac’s Ingenious Idea**

For a general qubit, two states labeled $|0\rangle$ and $|1\rangle$ associated with the possible values of some two-valued observable are chosen as the standard or ‘computational’ states—the reference states for computations. To remind you: the state $|0\rangle$ is just another qubit state like $|1\rangle$ and should not be confused with the number 0. For a photon, these states are usually the horizontal and vertical states of linear polarization in some standard direction, $|H\rangle$ and $|V\rangle$. For an electron, $|0\rangle$ and $|1\rangle$ usually represent the ‘up’ and ‘down’ states of spin in the direction $z$, for some standard direction chosen as the $z$ direction.

The qubit states $|0\rangle$ and $|1\rangle$ are represented by orthogonal vectors of unit length in a 2-dimensional state space. In the infinite-dimensional case, the state space is technically a Hilbert space (don’t worry—this doesn’t play any explicit role in the following). The term is also used for a finite-dimensional space like the space of qubit states. The states $|0\rangle$ and $|1\rangle$ define a coordinate system or ‘basis’ in the state space called the standard or computational basis, just as the unit vectors $\vec{x}$ and $\vec{y}$ in the directions of the $x$ and $y$ coordinates of a Cartesian coordinate system define the two orthogonal directions of a basis in the Cartesian plane. Any qubit state $|\psi\rangle$ represented by a unit vector at an angle $\theta$ to $|0\rangle$ can be represented as a linear combination or sum of its projections along $|0\rangle$ and $|1\rangle$ as $|\psi\rangle = \cos \theta |0\rangle + \sin \theta |1\rangle$, where the coefficients $\cos \theta$ and $\sin \theta$ represent the lengths of the projections of the unit state vector onto the directions defined by the unit length basis vectors $|0\rangle$ and $|1\rangle$. A linear combination or sum of states like $|\psi\rangle = \cos \theta |0\rangle + \sin \theta |1\rangle$ is referred to as a ‘superposition’ of the states $|0\rangle$ and $|1\rangle$. Any state can be expressed as a superposition of other states in an infinite number of ways, just as any vector can
be expressed as a linear combination of other vectors.

There is a difference between the quantum state space and the space of vectors in the Cartesian plane. To accommodate the probability relations between different polarization states, you need to consider linear combinations of unit vectors with coefficients that can be complex numbers.

If you are unfamiliar with complex numbers, here’s quick tutorial. A complex number is a number of the form \( c = a + ib \), where \( i = \sqrt{-1} \), so \( i^2 = -1 \). In effect, a new number \( i \) is defined as the square root of \(-1\): if you square \( i \), you get \(-1\). This gives you a whole new set of numbers to play with, the complex numbers, that are sums of ‘real’ numbers and multiples of real numbers by \( i \). The number \( a \) is referred to as the real part of \( c \), denoted by \( \text{Re}(c) \), and \( b \) as the imaginary part of \( c \), denoted by \( \text{Im}(c) \). The real numbers then become a subset of the complex numbers—the complex numbers with a zero imaginary part.

The ‘complex conjugate’ of \( c \) is defined as \( c^* = a - ib \). So the real part of \( c \) is \( \text{Re}(c) = \frac{1}{2}(c + c^*) \), because you get \( 2a \) when you add \( a + ib \) and \( a - ib \). The imaginary part of \( c \) is \( \text{Im}(c) = \frac{1}{2}(c^* - c) \), because subtracting \( a + ib \) from \( a - ib \) gives you \(-2ib \), which becomes \( 2b \) when you multiply by \( i \), since \( i \cdot i = -1 \). The product of \( c \) and \( c^* \) is a real number:

\[
cc^* = (a + ib)(a - ib) = a^2 + b^2
\]

(the product of \( ib \) and \(-ib \) is \(-i \cdot i \cdot b^2 = b^2 \)). This is taken as the square of the absolute value of the complex number \( c \) and denoted by \( |c|^2 \).

The only other feature of complex numbers relevant to the representation of quantum states here is the Euler relation:

\[
e^{i\theta} = \cos \theta + i \sin \theta
\]

where \( e \) is the base of the natural logarithm. If you have a hard time with the idea of an imaginary exponent, take the left hand side of the Euler equation as a shorthand expression for the right hand side, which is just a complex number. The relation follows from the expansion of \( e^{i\theta} \), \( \cos \theta \), and \( \sin \theta \) as infinite series. The complex number \( e^{i\theta} \) has absolute value 1. The complex conjugate of \( e^{i\theta} \) is \( e^{-i\theta} \), and \( e^{i\theta} e^{-i\theta} = 1 \). Alternatively, this follows because \( (\cos \theta + i \sin \theta)(\cos \theta - i \sin \theta) = \cos^2 \theta + \sin^2 \theta = 1 \).

In quantum mechanics, observables or physical quantities, like the polarization of a photon or the spin of an electron, are represented by operators that transform states represented by vectors to new states.

The neat thing about the Dirac notation is the clever way of representing operators. The projection operator \( P_{|0\rangle} \) onto the basis vector \( |0\rangle \) takes a general qubit state represented by a unit vector \( |\psi\rangle \) as input and produces as output the vector that
is the projection of $|\psi\rangle$ onto $|0\rangle$. So $|\psi\rangle$ can be expressed in terms of its projections onto the basis states $|0\rangle$ and $|1\rangle$ as a superposition, $|\psi\rangle = P_{|0\rangle}|\psi\rangle + P_{|1\rangle}|\psi\rangle$. This is just like expressing a vector $\vec{v}$ in terms of its projections $P_x \vec{v}$ and $P_y \vec{v}$ on the $x$ and $y$ axes of a Cartesian coordinate system as a linear combination $\vec{v} = P_x \vec{v} + P_y \vec{v}$.

(Here $P_{|0\rangle}|\psi\rangle = \cos \theta |0\rangle$ and $P_{|1\rangle}|\psi\rangle = \sin \theta |1\rangle$ because $|\psi\rangle$ is a unit vector.) In the Dirac notation, $P_{|0\rangle}$ is written as $|0\rangle\langle 0|$ and $P_{|1\rangle}$ as $|1\rangle\langle 1|$. Dirac calls a vector like $|0\rangle$ a ‘ket’ and a vector like $\langle 0|$ a ‘bra.’ The bra vector of $c|\psi\rangle$, for some complex number $c$, is $\langle \psi|c^*|$, where $c^*$ is the complex conjugate of $c$.

Here’s how the notation works: to obtain the output of the projection of a state vector $|\psi\rangle$ onto another state vector $|\phi\rangle$, apply the projection operator $|\phi\rangle\langle \phi|$ to $|\psi\rangle$. The output $\langle \phi|\phi|\psi\rangle$ is read as $\langle \phi|\phi|\psi\rangle$: the two vertical middle lines together are treated as a single vertical line. Dirac calls the expression $\langle \phi|\phi|$ a ‘bra(c)ket,’ hence the ‘bra,’ ‘ket’ terminology. It represents a number: the ‘length’ of the projection of $|\psi\rangle$ onto $|\phi\rangle$, or technically the ‘scalar product’ of $|\phi\rangle$ and $|\psi\rangle$. The term ‘scalar’ refers to a number rather than a vector, so a scalar product is a type of product of vectors that produces a number. The term ‘length’ is in quotes here because the number could be a complex number.

Since $\langle \phi|\psi\rangle$ is just a number, it can be moved to a position in front of the state $|\phi\rangle$, so:

$$\langle \phi|\phi|\psi\rangle = \langle \phi|\psi\rangle = (\langle \phi|\psi\rangle|\phi\rangle$$

Dirac’s neat idea is that re-grouping the expressions according to the underbraces allows you to see $\langle \phi|\phi|$ operating on $|\psi\rangle$ as a multiple of the unit vector $|\phi\rangle$, which is what you would expect after projecting $|\psi\rangle$ onto $|\phi\rangle$.

To sum up:

- The projection operator $|\phi\rangle\langle \phi|$ transforms a unit vector $|\psi\rangle$ to the vector $|\phi\rangle\langle \phi|\psi\rangle = \langle \phi|\psi\rangle|\phi\rangle$, which is the projection of $|\psi\rangle$ onto $|\phi\rangle$. In other words, a projection operator is a function that takes a vector as input and produces another vector as output, the projection of the original vector along a particular direction in the state space.

- The expression $|\phi\rangle\langle \phi|$ represents an operator, the projection operator onto $|\phi\rangle$. In general, an expression like $|\phi\rangle\langle \psi|$—note the $\langle \psi|$ here—represents an operator that transforms a vector $|\chi\rangle$ to a multiple of $|\phi\rangle$: $|\phi\rangle\langle \psi|\chi\rangle = |\phi\rangle\langle \psi|\chi\rangle = \langle \psi|\chi\rangle |\phi\rangle$.

- The expression $\langle \phi|\psi\rangle$ represents a number, the scalar product of $|\phi\rangle$ and $|\psi\rangle$, or the ‘length’ of the projection of $|\psi\rangle$ onto $|\phi\rangle$. In particular, $\langle 0|1\rangle = \langle 1|0\rangle = 0$. The length of the projection of $|1\rangle$ onto $|0\rangle$ or $|0\rangle$ onto $|1\rangle$ is 0,
because these are orthogonal vectors. Also, $\langle 0|0 \rangle = \langle 1|1 \rangle = 1$. The length of the projection of $|0\rangle$ onto itself, or $|1\rangle$ onto itself, is 1, because these are unit length vectors. For a non-unit vector like $|\psi\rangle = c|0\rangle$, for some complex number $c$, $\langle \psi|\psi \rangle = \langle 0|c^*c|0 \rangle = |c|^2(0|0 \rangle = |c|^2 = |\psi|^2$, where $|\psi|^2$ denotes the square of the absolute value of the ‘length’ of $|\psi\rangle$.

- In the Dirac notation, a state $|\psi\rangle$ is represented as the sum of its projections onto the computational basis states as $|0\rangle\langle 0|\psi \rangle + |1\rangle\langle 1|\psi \rangle = \langle 0|\psi\rangle|0\rangle + \langle 1|\psi\rangle|1\rangle$, where $\langle 0|\psi \rangle$ is the ‘length’ of the projection of $|\psi\rangle$ onto $|0\rangle$ and $\langle 1|\psi \rangle$ is the ‘length’ of the projection of $|\psi\rangle$ onto $|1\rangle$.

You might want to read this section a couple of times to get the hang of the notation, especially the difference between $\langle \phi|\psi \rangle$, which is a number or scalar, and $|\phi\rangle\langle \psi |$, which is an operator.

### The bottom line
- This section shows how quantum states are represented as vectors in a state space, in a commonly used notation invented by Dirac. The standard states for a qubit, also called the ‘computational’ states (the reference states for computations), are represented as $|0\rangle$ and $|1\rangle$ in the Dirac notation.
- Observables or physical quantities, like the polarization of a photon or the spin of an electron, are represented by operators that act on quantum states to transform them to new states.
- There’s also a brief tutorial on complex numbers in this section.

### The Pauli Operators

You can add operators like $\cdots \langle \cdots |$ with complex numbers as coefficients to form new operators, and you can multiply operators by real or complex numbers.

So take the operator $X = |0\rangle\langle 1| + |1\rangle\langle 0| = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ and the two unit vectors, $|x_+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ and $|x_-\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle$. These quantum state vectors are often denoted as $|+\rangle$ and $|−\rangle$. Here I want to consider operators $X$ and $Y$ and distinguish the state vectors associated with $X$ from the state vectors associated with $Y$, so I’ll write $|x_+\rangle$ and $|x_-\rangle$ rather than $|+\rangle$ and $|−\rangle$. The coefficients $\frac{1}{\sqrt{2}}$ and $−\frac{1}{\sqrt{2}}$ in $|x_+\rangle$ and $|x_-\rangle$ are just ‘normalizing’ factors that ensure that the vectors $|x_+\rangle$ and $|x_-\rangle$ are unit vectors, with length or ‘norm’ 1.
To find the length of a vector, you take the squares of the lengths of the orthogonal projections, add them, and take the square root. This is an application of Pythagoras’s theorem: the square of the length of the diagonal of a triangle is the sum of the squares of the lengths of the two orthogonal sides. The length of the diagonal here is the length of the vector, and the orthogonal sides of the triangle are the projections of the vector onto \(|0\rangle\) and \(|1\rangle\). The length of \(|x_+\rangle\), then, is the square root of \((\frac{1}{\sqrt{2}})^2 + (\frac{1}{\sqrt{2}})^2 = \frac{1}{2} + \frac{1}{2} = 1\), and the length of \(|x_-\rangle\) is the square root of \((\frac{1}{\sqrt{2}})^2 + (-\frac{1}{\sqrt{2}})^2 = \frac{1}{2} + \frac{1}{2} = 1\).

Alternatively, you can find the length of \(|x_+\rangle\) by calculating \(\langle x_+ | x_+ \rangle\), the length of the projection of the unit vector \(|x_+\rangle\) onto itself, and the length of \(|x_-\rangle\) by calculating \(\langle x_- | x_- \rangle\).

It’s easier to read a quantum state represented by a superposition of unit vectors with equal coefficients without the coefficients: \(|0\rangle + |1\rangle\) instead of \(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle\). In that case, the state is said to be ‘unnormalized,’ which is a technical way of saying that the state vector doesn’t have unit length (if the two coefficients are both 1, the length of the state vector is \(\sqrt{1 + 1} = \sqrt{2}\), by Pythagoras’s theorem). I’ll routinely leave out the coefficients when they are all equal and it’s clear how to put them back in to ‘normalize’ the state to unit length. (So they are left out in the expressions for the quantum states in the following paragraph.) If there are two terms in the superposition, the two equal coefficients are \(\frac{1}{\sqrt{2}}\). If there are three terms in the superposition, the three equal coefficients are \(\frac{1}{\sqrt{3}}\), and so on.

The unit vectors, \(|x_+\rangle = |0\rangle + |1\rangle\) and \(|x_-\rangle = |0\rangle - |1\rangle\) are uniquely associated with \(X\), in the sense that \(X|x_+\rangle = |x_+\rangle\) and \(X|x_-\rangle = -|x_-\rangle\). The result of operating on \(|x_+\rangle = |0\rangle + |1\rangle\) with the operator \(X = |0\rangle \langle 1| + |1\rangle \langle 0|\) is the sum of the results of operating on each of the components of this vector in turn.

To work this out, take the operation on the components of \(|x_+\rangle\) in order. The result of the first term \(|0\rangle \langle 1|\) in the operator \(X\) operating on the first component \(|0\rangle\) of \(|x_+\rangle\) is \(|0\rangle \langle 1|0\rangle = 0\), because \(\langle 1|0\rangle = 0\). Similarly, the result of the second term \(|1\rangle \langle 0|\) in \(X\) operating on the component \(|0\rangle\) is \(|1\rangle \langle 0|0\rangle = |1\rangle\), because \(\langle 0|0\rangle = 1\). So \(X\) operating on the first component \(|0\rangle\) of \(|x_+\rangle\) produces the vector \(|1\rangle\).

The first term \(|0\rangle \langle 1|\) in \(X\) operating on the second component \(|1\rangle\) of \(|x_+\rangle\) produces \(|0\rangle\), and the second term \(|1\rangle \langle 0|\) in \(X\) operating on the component \(|1\rangle\) produces 0. So \(X\) operating on the second component \(|1\rangle\) of \(|x_+\rangle\) produces \(|0\rangle\).

The net effect of \(X\) operating on \(|x_+\rangle = |0\rangle + |1\rangle\), then, is to produce \(|1\rangle + |0\rangle = |0\rangle + |1\rangle\), so \(X\) operating on \(|x_+\rangle\) reproduces \(|x_+\rangle\). Applied to \(|x_-\rangle = |0\rangle - |1\rangle\), the operator \(X\) produces \(|1\rangle - |0\rangle\), which is \(-|x_-\rangle = -(|0\rangle - |1\rangle\)).

The operator \(X\) operating on \(|x_+\rangle\) or \(|x_-\rangle\) leaves these vectors unchanged, except for multiplication by a factor of 1 or –1, respectively. This is not going to be the case for any other vectors in the qubit state space, because they are all repre-
sented by linear superpositions of these basis vectors, with non-zero coefficients. The vectors $|x_+\rangle$ and $|x_-\rangle$ are referred to as the ‘eigenvectors’ of the operator $X$, and 1 and $-1$ as the ‘eigenvalues’ of $X$ (from the German word ‘eigen,’ which in this sense means ‘characteristic of’).

A minor caveat is necessary here, because the previous comment is not quite correct. If $X$ operates on $-|x_+\rangle$ or $-|x_-\rangle$, the effect is also to leave these vectors unchanged, except for multiplication by a factor 1 or $-1$. In fact, the vectors $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$, for any $\theta$, are regarded as representing the same qubit state. This covers the case $-|\psi\rangle$, because $e^{i\pi} = \cos \pi + i \sin \pi = -1$ (the cosine of $\pi$ is $-1$, and the sine of $\pi$ is 0), and the case $i|\psi\rangle$, because $e^{i\pi/2} = \cos \pi/2 + i \sin \pi/2 = i$ (the cosine of $\pi/2$ is 0 and the sine of $\pi/2$ is 1). The set of vectors \{e^{i\theta}|\psi\rangle\}, for all $\theta$, defines a 1-dimensional subspace or ‘ray.’ Any unit vector in the ray can be taken as representing the ray. The factor $e^{i\theta}$, which has absolute value 1, is referred to as a ‘phase,’ and identifies different vectors in the ray. So, to be precise, a qubit state is represented by a ray or 1-dimensional subspace in the state space, not a vector, but for simplicity you can take a unit vector in the subspace to specify the state. Alternatively, take the projection operator onto the 1-dimensional subspace to represent the state.

While the global phase makes no difference to the probabilities defined by a quantum state vector—$e^{i\theta}|\psi\rangle$ assigns the same probabilities to the possible values of observables as the state vector $|\psi\rangle$—relative phase does make a difference. In the section Relative Phase I show that the state vectors $|x_+\rangle = |0\rangle + |1\rangle$ and $|x_\theta\rangle = |0\rangle + e^{i\theta}|1\rangle$ are different qubit states and assign different probabilities to the possible outcomes of an $X$-measurement.

The eigenvectors of $X$ are orthogonal: their scalar product $\langle x_+ | x_- \rangle$ or $\langle x_- | x_+ \rangle$ is 0. This is always the case for the different eigenvectors of an operator. To compute an expression like $\langle x_+ | x_- \rangle$ when $|x_+\rangle$ and $|x_-\rangle$ are linear superpositions, express $\langle x_+ | x_- \rangle$ explicitly as $\langle 0 \rangle + \langle 1 \rangle$ and take the scalar products of the terms in this expression with the terms in the expression $|0\rangle - |1\rangle$ for $|x_-\rangle$, in order, adding or subtracting the results as appropriate depending on whether the coefficients are 1 or $-1$. The scalar products of $\langle 0 \rangle + \langle 1 \rangle$ and $|0\rangle - |1\rangle$ term by term are $\langle 0 | 0 \rangle + \langle 1 | 0 \rangle - \langle 0 | 1 \rangle - \langle 1 | 1 \rangle$, so $\langle x_+ | x_- \rangle = \frac{1}{2} (|0\rangle|0\rangle + |1\rangle|0\rangle - |0\rangle|1\rangle - |1\rangle|1\rangle) = \frac{1}{2} (1 + 0 - 0 - 1) = 0$.

Similarly, you can check that the operator $Y = -i|0\rangle\langle 1 | + i|1\rangle\langle 0 |$ has the eigenvectors $|y_+\rangle = |0\rangle + i|1\rangle$ and $|y_-\rangle = |0\rangle - i|1\rangle$ with eigenvalues $\pm 1$, and the computational basis states $|0\rangle$ and $|1\rangle$ are the eigenvectors of the operator $Z = |0\rangle\langle 0 | - |1\rangle\langle 1 |$, so $Z|0\rangle = |0\rangle$ and $Z|1\rangle = -|1\rangle$.

The three operators $X, Y, Z$ are called the Pauli operators. They represent three qubit observables, each with possible values $\pm 1$, the eigenvalues of the observables. (I’ll use the same symbol to represent the observable and the corresponding
operator.) The two eigenvectors of each observable define a basis associated with the observable. As I mentioned before, the convention is to take the eigenvectors of $Z$ as the standard or computational basis, and to represent the eigenvector $|z_+\rangle$ as $|0\rangle$ and the eigenvector $|z_-\rangle$ as $|1\rangle$. Then the eigenvector $|0\rangle$ corresponds to the eigenvalue 1 and the eigenvector $|1\rangle$ corresponds to the eigenvalue $-1$, which might be a little confusing at first.

To emphasize again: the eigenvalues are just convenient labels for the two possible values of the observables. It’s convenient for computational and information-theoretic purposes to represent the eigenvalues of a conventionally chosen ‘computational basis’ observable of a qubit as the bits, 0 or 1, but for other purposes it’s convenient to take the eigenvalues of spin observables or polarization observables as $\pm 1$. What’s physically relevant is the relation between different observables, expressed by the geometrical relation between their associated eigenvectors, the different basis states corresponding to the observables.

Here’s a summary:

- $X = |0\rangle\langle 1| + |1\rangle\langle 0|$. Eigenvectors: $|x_+\rangle = |0\rangle + |1\rangle$, $|x_-\rangle = |0\rangle - |1\rangle$
- $Y = -i|0\rangle\langle 1| + i|1\rangle\langle 0|$. Eigenvectors: $|y_+\rangle = |0\rangle + i|1\rangle$, $|y_-\rangle = |0\rangle - i|1\rangle$
- $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$. Eigenvectors: $|z_+\rangle = |0\rangle$, $|z_-\rangle = |1\rangle$

Finally, the identity operator, $I$, can be expressed as $I = |0\rangle\langle 0| + |1\rangle\langle 1|$. The identity is like the number 1, which leaves any number the same when you multiply by 1. So, for example, $I\cdot Z = (|0\rangle\langle 0| + |1\rangle\langle 1|)\cdot(\langle 0|\langle 0| - \langle 1|\langle 1|) = |0\rangle\langle 0| - \langle 1|\langle 1| = Z$, and $I|\psi\rangle = |\psi\rangle$ for any state $|\psi\rangle$. You can check that $X^2 = Y^2 = Z^2 = I$ (see the section Quantum Dynamics for an explicit calculation).

The identity operator is the sum of projection operators onto two orthogonal directions in the 2-dimensional qubit space, so it could equally well be expressed as $I = |x_+\rangle\langle x_+| + |x_-\rangle\langle x_-|$ or as $I = |y_+\rangle\langle y_+| + |y_-\rangle\langle y_-|$, or in terms of any other basis in the state space.

The notation $X, Y, Z$ for these operators suggests an orthogonal coordinate system, with $x, y, z$ axes. In fact, for electrons, $X, Y, Z$ represent spin observables for three orthogonal directions of spin: spin in the direction $x$, spin in the direction $y$, and spin in the direction $z$. For photons, the eigenstates of $Z$ are the horizontal and vertical linear polarization states for some standard direction $\vec{z}$: $|H\rangle = |0\rangle, |V\rangle = |1\rangle$, the eigenstates of $X$ are the two orthogonal states of diagonal polarization at an angle 45° to $\vec{z}$: $|\langle \rangle\rangle = |x_+\rangle, |\langle \rangle\rangle = |x_-\rangle$, and the eigenstates of $Y$ are the two states of right and left circular polarization: $|\langle \rangle\rangle = |y_+\rangle, |\langle \rangle\rangle = |y_-\rangle$.

You might wonder why the eigenstates of $Y$ involve complex numbers while the eigenstates of $X$ and $Z$ do not. The short answer is that you can’t fit the
representation of spins on the three-dimensional Bloch sphere or polarizations on the 3-dimensional Poincaré sphere onto a two-dimensional qubit space without using complex numbers. So some observables and eigenstates will need to involve complex numbers in their expressions, and the choice ultimately depends on getting the experimental probabilities right.

The representation $Z = |z_+\rangle\langle z_+| - |z_-\rangle\langle z_-|$, in which the operator is represented as a sum of the projection operators onto the eigenstates of the operator, with the eigenvalues as coefficients, is called the ‘spectral representation’ of the operator. The ‘spectrum’ of an operator or the corresponding observable is the set of eigenvalues, the possible values of the observable, in this case labeled 1 for the eigenvalue that is the coefficient of $|z_+\rangle\langle z_+|$ and $-1$ for the eigenvalue that is the coefficient of $|z_-\rangle\langle z_-|$. (Alternatively, you could write $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$ to make explicit that $|z_+\rangle$ and $|z_-\rangle$ are the computational basis eigenvectors $|0\rangle$ and $|1\rangle$.)

You can check that the spectral representations of $X$ and $Y$ are $X = |x_+\rangle\langle x_+| - |x_-\rangle\langle x_-|$ and $Y = |y_+\rangle\langle y_+| - |y_-\rangle\langle y_-|$ by substituting the expressions for the eigenvectors in the computational basis to obtain the representations of $X$ and $Y$ in the computational basis: $X = |0\rangle\langle 1| + |1\rangle\langle 0|$ and $Y = -i|0\rangle\langle 1| + i|1\rangle\langle 0|$. So, to be clear, $|x_+\rangle\langle x_+| - |x_-\rangle\langle x_-| = |0\rangle\langle 1| + |1\rangle\langle 0|$ and $|y_+\rangle\langle y_+| - |y_-\rangle\langle y_-| = -i|0\rangle\langle 1| + i|1\rangle\langle 0|$.

Not all operators in the state space represent observables. Operators that represent observables and can be associated with eigenvectors and associated real-valued eigenvalues in this way form a special class of operators in the state space: the ‘self-adjoint’ or ‘Hermitian’ operators. The adjoint of an operator is obtained by replacing $i$ with $-i$ (so replacing complex numbers with their complex conjugates) and transposing bras and kets (so replacing $|\psi\rangle\langle \phi|$ with $|\phi\rangle\langle \psi|$). In the section States and Operators as Matrices I’ll show that operators can be represented by matrices. Transposing bras and kets of an operator in the Dirac notation is equivalent to transposing rows and columns of the corresponding matrix, which produces the ‘transpose’ of the matrix.

The adjoint of $X = |0\rangle\langle 1| + |1\rangle\langle 0|$ is $X$, because the coefficients are real numbers, and exchanging bras and kets in the two terms changes the first term to the second term, and the second term to the first term. Since the terms are added, the operator is unchanged. Similarly, the adjoint of $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$ is $Z$, because all the coefficients are real numbers, and switching bras and kets leaves the operator unchanged. In the case of $Y$, denoting the adjoint by $Y^*$, you get $Y^* = (-i|0\rangle\langle 1| + i|1\rangle\langle 0|)^* = i|1\rangle\langle 0| - i|0\rangle\langle 1| = Y$. So the three Pauli operators are all self-adjoint.

To sum up: The quantum states of a qubit are represented by unit vectors (more properly by 1-dimensional subspaces or rays) on a complex vector space or linear
space (‘linear’ referring to the way vectors are composed by vector addition to produce new vectors; ‘complex’ because the coefficients multiplying the vectors are real or complex numbers, and real numbers are just a special case of complex numbers). Observables are represented by self-adjoint operators on the state space. Each such operator is associated with a basis of orthogonal unit vectors in the state space, with each vector in the basis corresponding to one of the two possible binary values of the observable, representing the possible outcomes of a measurement of the observable.

This is the case for pure quantum states. There are also mixed states representing mixtures: probability distributions or convex combinations of pure states. For example, a qubit could be prepared in such a way that it is either in the state \( |0\rangle \) with probability \( p_0 \) or in the state \( |1\rangle \) with probability \( p_1 \), in which case the qubit is said to be in a mixed state. I’ll show how to represent mixed states by operators in the section Mixed States later in this Supplement.

It turns out that in quantum mechanics, unlike in classical theories, the representation of a mixed state as a particular mixture of pure states is not unique. The maximally mixed state of a qubit represented as an equal weight mixture of the pure states \( |0\rangle \) and \( |1\rangle \) with \( p_1 = p_2 = 1/2 \) is equivalent to an equal weight mixture of the pure states \( |x_+\rangle \) and \( |x_-\rangle \). Measuring \( Z \) or \( X \) or any other observable on either of these mixtures will produce the same distribution of measurement outcomes. So there’s no way to distinguish a qubit that has been prepared in such a way that it is either in the state \( |0\rangle \) or in the state \( |1\rangle \) with probability 1/2 from a qubit that has been prepared in such a way that it is either in the state \( |x_+\rangle \) or in the state \( |x_-\rangle \) with probability 1/2. In fact, a maximally mixed state is equivalent to any equal weight mixture of orthogonal pure states, as well as an infinite set of mixtures of nonorthogonal pure states.
The bottom line

- The quantum states of a qubit are represented by unit vectors, and the observables by operators. A measurement of a qubit observable has two possible outcomes, represented by the two ‘eigenvalues’ of the observable and two corresponding orthogonal state vectors, the ‘eigenvectors’ of the observable. The eigenvectors define a coordinate system or basis in the state space of the qubit.

- The Pauli observables, denoted by $X, Y, Z$, are three qubit observables that play a special role throughout the book. For electrons, $X, Y, Z$ represent spin observables for three orthogonal directions of spin: spin in the direction $x$, spin in the direction $y$, and spin in the direction $z$. For photons, the eigenstates of $Z$ are the horizontal and vertical linear polarization states for some standard direction $z$, the eigenstates of $X$ are the two orthogonal states of diagonal polarization at an angle 45° to $z$, and the eigenstates of $Y$ are the two states of right and left circular polarization. The convention is to take the eigenbasis of $Z$ as the standard or computational basis, and to represent the eigenvectors as $|0\rangle$ and $|1\rangle$.

- There are also mixed states representing mixtures: probability distributions of pure states. For example, a qubit could be prepared in such a way that it is either in the state $|0\rangle$ with probability $p_0$ or in the state $|1\rangle$ with probability $p_1$. I show how to represent mixed states by operators in the section Mixed States.

- In the Bananaworld chapter, peeling $S$ corresponds to measuring the $Z$ observable of a qubit, peeling $T$ corresponds to measuring the $X$ observable, and the tastes 0 and 1 correspond to the two possible values of $Z$ and $X$. In the chapter Quantum Magic, the correspondence between quantum observables and peelings is different.

The Born Rule

The relation of states and observables to what we see in quantum phenomena is given by the Born probability rule. If a qubit in the pure state $|\psi\rangle$ is subjected to a measurement, say an $X$-measurement, the probabilities of the two possible outcomes, $\pm 1$ corresponding to the eigenstates $|x_+\rangle, |x_-\rangle$, are given by the squares of the absolute values of the ‘lengths’ of the projections of $|\psi\rangle$ onto the eigenstates.
of $X$ (the quotes are there because this ‘length’ could be a complex number):

$$\text{prob}_\psi (1) = |P_{x+}|\psi \rangle |^2 = |\langle x_+ |\psi \rangle |^2$$

$$\text{prob}_\psi (-1) = |P_{x-}|\psi \rangle |^2 = |\langle x_- |\psi \rangle |^2$$

After the measurement, the qubit will either be in the state $|x_+\rangle$ or in the state $|x_-\rangle$, depending on the outcome. For example, a photon subjected to a measurement of linear polarization in a particular direction $z$ by passing it through a polarizing beamsplitter will exit the beamsplitter in one of two states of linear polarization: it will either be linearly polarized in the $z$ direction, or in a direction orthogonal to $z$. Although it’s customary, it’s a bit odd to refer to an $X$-measurement of a qubit in the state $|\psi \rangle$ as a measurement to check whether the qubit is in the state $|x_+\rangle$ or the state $|x_-\rangle$, since a qubit in any state, if subjected to an $X$-measurement, is forced to ‘choose’ between the two states $|x_+\rangle$ and $|x_-\rangle$. A measurement of a qubit, unlike a measurement in a classical theory, is a process that induces the qubit to make an intrinsically random transition to one of two alternative states, as I showed in Chapter ??, Really Random.

Another way of expressing the Born rule is in terms of the expectation value of an observable in a quantum state. The expectation value or average value of $X$ in the pure state $|\psi \rangle$, denoted by $\langle X \rangle_\psi$, is the sum of the possible values $\pm 1$ of $X$ weighted by the probabilities $p_+ = |P_{x+}|\psi \rangle |^2$ for 1 and $p_- = |P_{x-}|\psi \rangle |^2$ for $-1$ in the state $|\psi \rangle$: $\langle X \rangle_\psi = p_+ - p_-$. There is a concise way of expressing the expectation value in the Dirac notation. The projection operator $P_{x+}$ is self-adjoint (this is obvious because of its form: $P_{x+} = |x_+\rangle \langle x_+|$) and ‘idempotent’: $P_{x+} \cdot P_{x+} = P_{x+}$ (projecting twice is the same as projecting once). So, just as $|\psi \rangle |^2 = \langle \psi \| \psi \rangle$ is the square of the absolute value of the ‘length’ of the vector $|\psi \rangle$, so $|P_{x+}|\psi \rangle |^2$ can be expressed as $\langle \psi \| P_{x+} \cdot P_{x+} |\psi \rangle = \langle \psi \| P_{x+} |\psi \rangle$. Similarly, $|P_{x-}|\psi \rangle |^2 = \langle \psi \| P_{x-} |\psi \rangle$. So

$$\langle X \rangle_\psi = \langle \psi \| P_{x+} |\psi \rangle - \langle \psi \| P_{x-} |\psi \rangle$$

$$= \langle \psi \| P_{x+} - P_{x-} |\psi \rangle$$

$$= \langle \psi \| X |\psi \rangle$$

because $|P_{x+} - P_{x-}| = X$ in the spectral representation. This is a general expression for the expectation value of any observable $O$ in a pure quantum state $|\psi \rangle$: $\langle O \rangle_\psi = \langle \psi \| O |\psi \rangle$.

The Born rule has an extension for mixed states. See the section Mixed States.

The Pauli operators are ‘mutually unbiased.’ The absolute value of the ‘length’ of the projection of any eigenvector in one basis onto any eigenvector in another basis is the same: $1/\sqrt{2}$. This is clear from the expressions for the $X$-eigenvectors
and the $Y$-eigenvectors in terms of the $Z$-eigenvectors $|0\rangle$ and $|1\rangle$, because the coefficients are $1/\sqrt{2}$ for $|0\rangle$ and $\pm 1/\sqrt{2}$ or $\pm i/\sqrt{2}$ for $|1\rangle$, and these all have absolute value $1/\sqrt{2}$. To see that this is so for the projections of $Y$-eigenvectors onto $X$-eigenvectors, or $Z$-eigenvectors onto $Y$-eigenvectors, you could express $Y$-eigenvectors in terms of $X$-eigenvectors, or $Z$-eigenvectors in terms of $Y$-eigenvectors, which would be rather tedious, or simply observe that the relations between $X, Y, Z$ are unchanged if you permute them cyclically. For example, the relations between these observables are unchanged if you re-label $X$ as $Y$ and $Y$ as $Z$ and $Z$ as $X$.

So, by the Born rule, if a qubit in an eigenstate of one Pauli observable is subjected to a measurement of a different Pauli observable, the probability is $1/2$—the square of $\pm 1/\sqrt{2}$ or the absolute value of the square of $\pm i\sqrt{2}$—that the qubit will be detected as having the eigenvalue $1$, and $1/2$ that the qubit will be detected as having the eigenvalue $-1$. For example, if a photon in any state represented by a vector in one of the Pauli bases hits a polarizing filter associated with any other Pauli basis, it has an even chance of getting through and an even chance of being blocked.

**The bottom line**

- The relation of states and observables to what we see in quantum phenomena is given by the Born probability rule. If a qubit in the pure state $|\psi\rangle$ is subjected to a measurement, say an $X$-measurement, the probabilities of the two possible outcomes, $\pm 1$ corresponding to the eigenstates $|x_+\rangle$, $|x_-\rangle$, are given by the squares of the absolute values of the ‘lengths’ of the projections of $|\psi\rangle$ onto the eigenstates of $X$ (the quotes are there because this ‘length’ could be a complex number).

- Another way of expressing the Born rule is in terms of the expectation value of an observable in a quantum state. There is a concise way of expressing the expectation value in the Dirac notation.

- The Pauli operators are ‘mutually unbiased.’ By the Born rule, if a qubit in an eigenstate of one Pauli observable is subjected to a measurement associated with a different Pauli observable, the probability is $1/2$ of the qubit being detected as having the eigenvalue $1$ and $1/2$ of the qubit being detected as having the eigenvalue $-1$. For example, if a photon in any state represented by a vector in one of the Pauli bases hits a polarizing filter associated with any other Pauli basis, it has an even chance of getting through and an even chance of being blocked.
Noncommutativity and Uncertainty

Operators can be multiplied and the multiplication is associative: \( A(B + C) = AB + AC \), for any operators \( A, B, C \). As an example, take \( A = |0\rangle\langle 0|, B = |0\rangle\langle 1| \), and \( C = |1\rangle\langle 0| \). Then \( |0\rangle\langle 0| \cdot (|0\rangle\langle 1| + |1\rangle\langle 0|) = |0\rangle\langle 0|\langle 1| + |0\rangle\langle 0|\langle 0| = |0\rangle\langle 1| \), because \( |0\rangle\langle 0| \) sandwiched between the vectors \( |0\rangle \) and \( |1\rangle \) is just the number \( \langle 0|0\rangle = 1 \), and \( |0\rangle\langle 1| \) sandwiched between the vectors \( |0\rangle \) and \( |0\rangle \) is the number \( \langle 0|1\rangle = 0 \).

The operation of multiplying two operators, \( A \) and \( B \), needn’t commute. It could be that \( AB \) operating on a vector doesn’t produce the same output as \( BA \) operating on the vector—the order of applying the operations, first \( B \) and then \( A \), or first \( A \) and then \( B \), could make a difference. For example, \( XY = (|0\rangle\langle 1| + |1\rangle\langle 0|) \cdot (-i|0\rangle\langle 1| + i|1\rangle\langle 0|) = i|0\rangle\langle 0| - i|1\rangle\langle 1| = iZ \), but \( YX = (-i|0\rangle\langle 1| + i|1\rangle\langle 0|) \cdot (|0\rangle\langle 1| + |1\rangle\langle 0|) = -i|0\rangle\langle 0| + i|1\rangle\langle 1| = -iZ \). So \( XY - YX = 2iZ \).

The difference \( XY - YX \), called the ‘commutator’ of \( X \) and \( Y \), is denoted by \( [X, Y] = 2iZ \). Similarly, you can check that \( YZ = iX, ZY = -iX \), so \( [Y, Z] = 2iX \), and \( ZX = iY, XZ = -iY, \) so \( [Z, X] = 2iY \). The commutation relations are the same for the triples \( X, Y, Z \) and \( Y, Z, X \) and \( Z, X, Y \), so for cyclic permutations of \( X, Y, Z \).

Operators associated with different bases in the state space don’t commute. Operating on a state vector in the state space with two of these operators in succession (represented by multiplying the two operators in a certain order) is not generally the same thing as operating on the state vector with the two operators in the reverse order: the output vector of the combined operation is different in the two cases.

Heisenberg’s uncertainty or indeterminacy principle is a consequence of noncommutativity. In Heisenberg’s original version, the noncommuting operators represent position and momentum, usually denoted by \( \hat{x} \) for position and \( \hat{p} \) for momentum. In this case the commutator turns out to be a multiple of the identity operator \( I \): \( [\hat{x}, \hat{p}] = i\hbar I \), or \( [\hat{x}, \hat{p}] = iI \) in units in which \( \hbar \), Planck’s constant divided by \( 2\pi \), is equal to 1, which I’ve implicitly assumed throughout the book. From this relation between the position and momentum operators, you get a reciprocal relation between the extent to which it is possible to fix the range of possible position values and the range of possible momentum values for a system in any quantum state.

There is a similar uncertainty principle for the Pauli observables of a qubit. If a qubit is in a particular eigenstate of \( X, Y \), or \( Z \), there’s an even probability of yielding the outcome \( \pm 1 \) in a measurement of any other Pauli observable. The values of the different Pauli observables can’t be specified more precisely than this: if one Pauli observable has a definite value, the values of the other Pauli observables are completely indefinite.
Operators associated with different bases in the state space don’t commute. Operating on a state vector in the state space with two of these operators in succession (represented by multiplying the two operators in a certain order) is generally not the same thing as operating on the state vector with the two operators in the reverse order: the output vector of the combined operation is different in the two cases.

Heisenberg’s uncertainty or indeterminacy principle is a consequence of noncommutativity. In Heisenberg’s original version, the noncommuting operators represent position and momentum. There is a similar uncertainty principle for the Pauli observables of a qubit: if a qubit is in a particular eigenstate of \(X\), \(Y\), or \(Z\), there’s an even probability of yielding the outcome \(\pm 1\) in a measurement of any other Pauli observable. The values of the different Pauli observables can’t be specified more precisely than this.

Entanglement

If two qubits, \(A\) and \(B\), are separately in states \(|a⟩\) and \(|b⟩\), the state of the combined system is represented as \(|a⟩ \otimes |b⟩\), where the symbol \(\otimes\) represents a product called a tensor product. Technically, the states \(|a⟩\) and \(|b⟩\) are represented by vectors in the two-dimensional state spaces of the separate systems \(A\) and \(B\), and the state of the composite system, \(|a⟩ \otimes |b⟩\), is a vector in a new state space, the ‘tensor product’ of the two state spaces, which is a four-dimensional space. The tensor product space is constructed as the state space consisting of all product vectors of the form \(|a⟩ \otimes |b⟩\) (the Cartesian product of the two state spaces, the set of all ordered pairs of states from each of the state spaces), together with superpositions like \(c|a_1⟩ \otimes |b_1⟩ + d|a_2⟩ \otimes |b_2⟩\), where \(|a_1⟩\), \(|a_2⟩\) are two states of \(A\) and \(|b_1⟩\), \(|b_2⟩\) are two states of \(B\), and \(c\) and \(d\) are complex numbers.

It’s often easier to read such expressions without the tensor product symbol \(\otimes\), so unless it’s needed for clarity or to avoid ambiguity, it’s usual to write \(c|a_1⟩|b_1⟩ + d|a_2⟩|b_2⟩\) and even, for short, \(c|a_1b_1⟩ + d|a_2b_2⟩\).

Some linear superpositions like \(|0⟩|1⟩ + |1⟩|0⟩\) are really product states in disguise. This state is just \((|0⟩ + |1⟩) \otimes |1⟩ = |+⟩|1⟩\). So it’s a tensor product of the state \(|+⟩\) for qubit \(A\) and the state \(|1⟩\) for qubit \(B\). The two qubits are said to be separable, and measurements on the qubits are uncorrelated.

Other linear superpositions in the tensor product space, like: \(|ψ^−⟩ = |0⟩|1⟩ − |1⟩|0⟩\)
\[ |1\rangle|0\rangle \] cannot be expressed as product states. Such two-qubit pure states are referred to as ‘entangled’ pure states. An entangled two-qubit pure state is non-separable and cannot be expressed as a product state of two one-qubit pure states. Measurement outcomes on the two qubits are correlated. In the case of \[ |\psi^-\rangle \], measurement outcomes are perfectly anticorrelated (oppositely correlated) if the same observable is measured on both qubits. A mixed state of a bipartite system is entangled if it can’t be expressed as a mixture or probability distribution of product states of the two systems. You can have entangled states of any number of qubits or qutrits or any sort of quantum system.

Here’s how to see that \[ |\psi^-\rangle \] can’t be expressed as a product state. If \[ |\psi^-\rangle \] could be expressed as a product state \( (a|0\rangle + b|1\rangle) \otimes (c|0\rangle + d|1\rangle) = ac|0\rangle|0\rangle + bd|1\rangle|1\rangle + ad|0\rangle|1\rangle + bc|1\rangle|0\rangle \), the complex numbers \( a, b, c, d \) would have to satisfy the conditions \( ac = 1, bd = -1, ad = 0, bc = 0 \). If \( ad = 0 \), then at least one of \( a \) or \( d \) would have to be zero. If \( a = 0 \) then \( ac = 0 \), which is inconsistent with the condition \( ac = 1 \). If \( d = 0 \), then \( bd = 0 \), which is inconsistent with the condition \( bd = -1 \). So the conditions can’t be satisfied.

The entangled state \[ |\psi^-\rangle \] is a rather special state: it takes the same form for any basis (possibly up to multiplication by a global phase \( e^{i\theta} \)). For example:

\[ |0\rangle|1\rangle - |1\rangle|0\rangle = |+\rangle|\rangle - |\rangle|+\rangle \]

where \( |+\rangle, |\rangle \) are the \( X \)-eigenstates, and similarly for any other basis. What the entangled state \[ |\psi^-\rangle \] expresses is perfect anticorrelation. If Alice and Bob share two qubits in this entangled state, and Alice measures any observable of her qubit and Bob measures the same observable of his qubit, their measurement outcomes will be anticorrelated. (Reminder: the missing coefficients are both \( 1/\sqrt{2} \) here to ‘normalize’ the state to unit length.) If Alice and Bob both measure \( Z \), the probability is \( 1/2 \) (the square of \( 1/\sqrt{2} \)) that Alice will find the state \( |0\rangle \) and Bob will find the state \( |1\rangle \), and \( 1/2 \) that Alice will find the state \( |1\rangle \) and Bob will find the state \( |0\rangle \). Similarly, if Alice and Bob measure \( X \), there is an even probability that Alice will find the state \( |+\rangle \) and Bob the state \( |\rangle \), or that Alice will find the state \( |\rangle \) and Bob the state \( |+\rangle \), and similarly for any other qubit observable.

What if Alice measures \( Z \) and Bob does nothing? After the measurement, Alice’s qubit is in one of the eigenstates of \( Z \), \( |0\rangle \) or \( |1\rangle \). In the four-dimensional two-qubit state space, each outcome is associated with one of two orthogonal two-dimensional subspaces or planes, defined by the projection operators \( |0\rangle\langle 0|_A \otimes I_B \) and \( |1\rangle\langle 1|_A \otimes I_B \). Here \( I_B \) is the identity operator on Bob’s two-dimensional qubit space: operating on any vector in the space, the identity operator reproduces the vector. (I use subscripts here to distinguish Alice operators from Bob operators.)

According to the Born rule, if the state is \[ |\psi^-\rangle \], the probability of a particular outcome (if Alice measures \( Z \) and Bob does nothing) is the square of the absolute
value of the length of the projection of $|\psi^-\rangle$ onto the plane in the two-qubit space associated with the outcome. The projections onto the planes associated with the two possible measurement outcomes are $|0\rangle\langle 0|_A \otimes I_B \cdot |\psi^-\rangle$ and $|1\rangle\langle 1|_A \otimes I_B \cdot |\psi^-\rangle$. (The ‘·’ here is introduced to visually separate the projection operator from the state for clarity.) What are the lengths of these projections?

To work out the length of the projection $|0\rangle\langle 0|_A \otimes I_B \cdot |\psi^-\rangle$, write it out as $|0\rangle\langle 0|_A \otimes I_B \cdot (\frac{1}{\sqrt{2}}|0\rangle_1 B - \frac{1}{\sqrt{2}}|1\rangle_1 A)|0\rangle_B$, with the coefficients explicit. The $A$-operator acts on vectors in the $A$-space, and the $B$-operator acts on vectors in the $B$-space. Consider the first term in the state $\frac{1}{\sqrt{2}}|0\rangle_1 A|1\rangle_B - \frac{1}{\sqrt{2}}|1\rangle_1 A|0\rangle_B$. The identity $I_B$ operating on the vector $|1\rangle_B$ reproduces the vector $|1\rangle_B$, and $|0\rangle\langle 0|_A$ operating on $|0\rangle_A$ produces $|0\rangle_A$ (because $\langle 0|0\rangle_A = 1$). So the operator $|0\rangle\langle 0|_A \otimes I_B$ acting on the first term in the state $\frac{1}{\sqrt{2}}|0\rangle_1 A|1\rangle_B - \frac{1}{\sqrt{2}}|1\rangle_1 A|0\rangle_B$ produces the vector $\frac{1}{\sqrt{2}}|0\rangle_A |1\rangle_B$. Since the operator $|0\rangle\langle 0|_A \otimes I_B$ acting on the second term produces 0 (because $\langle 0|1\rangle_A = 0$), the projection is the vector $\frac{1}{\sqrt{2}}|0\rangle_A |1\rangle_B$, which has length $1/\sqrt{2}$ (because $|0\rangle_A$ and $|1\rangle_B$ are both unit vectors). The probability is the square of this length, which is $1/2$. A similar calculation shows that the probability for the projection operator $|1\rangle\langle 1|_A \otimes I_B \cdot |\psi^-\rangle$ is also $1/2$.

There are three other states like $|\psi^-\rangle$ that form a basis with $|\psi^-\rangle$, called the ‘Bell basis,’ in the two-qubit state space:

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}|0\rangle|0\rangle + \frac{1}{\sqrt{2}}|1\rangle|1\rangle$$
$$|\phi^-\rangle = \frac{1}{\sqrt{2}}|0\rangle|0\rangle - \frac{1}{\sqrt{2}}|1\rangle|1\rangle$$
$$|\psi^+\rangle = \frac{1}{\sqrt{2}}|0\rangle|1\rangle + \frac{1}{\sqrt{2}}|1\rangle|0\rangle$$
$$|\psi^-\rangle = \frac{1}{\sqrt{2}}|0\rangle|1\rangle - \frac{1}{\sqrt{2}}|1\rangle|0\rangle$$

As you can check, these states are mutually orthogonal: the scalar product of any two Bell states is zero. The correlations defined by $|\psi^-\rangle$ are the anticorrelations in the Bohm version of the Einstein–Podolsky–Rosen argument. The correlations of the state $|\phi^+\rangle$ are the correlations of Einstein–Podolsky–Rosen bananas. Peeling $S$ corresponds to measuring $Z$, and peeling $T$ corresponds to measuring $X$. The two tastes, 0 or 1, correspond to the eigenvalues 1 and $-1$. Like the state $|\psi^-\rangle$, the state $|\phi^+\rangle$ takes the same form in the $X$-basis and the $Z$-basis (but, in general, for Bell states other than $|\psi^-\rangle$, transforming between bases takes you from one Bell state to another Bell state). The four mutually orthogonal product states $|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle$ form an alternative computational basis in the two-qubit state space.
Bell’s insight was to see that, while the Einstein–Podolsky–Rosen correlations can be simulated with local resources (which is to say that there is a common cause explanation of these correlations), this is not the case for the probabilistic correlations of a Bell state if Alice and Bob measure different observables. For a particular choice of observables, it’s possible to reach the Tsirelson bound and produce correlations that are as far as is physically possible from correlations that can be simulated with local resources, and as close as is physically possible to the Popescu–Rohrlich correlations. So in a simulation game where Alice and Bob are required to simulate the correlations of Popescu–Rohrlich bananas, they can win the game with an optimal probability $\frac{1}{2}(1 + \frac{1}{\sqrt{2}}) \approx 0.85$ if they are allowed unlimited access to shared entangled quantum states. This is as close as you can get with quantum resources to the maximum probability 1 for a perfect simulation, and closer than the optimal probability .75 with local resources. They each measure specifically chosen observables to the prompts in the game and respond with the measurement outcomes.

**The bottom line**

- Some linear superpositions of two-qubit states like $|\psi^-\rangle = |0\rangle|1\rangle - |1\rangle|0\rangle$ are non-separable and cannot be expressed as a product state of two one-qubit pure states. Such two-qubit pure states are referred to as 'entangled' pure states. A mixed state of a bipartite system is entangled if it can’t be expressed as a mixture or probability distribution of product states of the two systems.

- Measurement outcomes on two entangled qubits are correlated. In the case of $|\psi^-\rangle$, measurement outcomes are perfectly anticorrelated if the same observable is measured on both qubits.

- The state $|\psi^-\rangle$ is one of four so-called Bell states, $|\psi^-\rangle, |\psi^+\rangle, |\phi^-\rangle, |\phi^+\rangle$, that form a basis in the two-qubit state space. The correlations defined by $|\psi^-\rangle$ are the anticorrelations in the Bohm version of the Einstein–Podolsky–Rosen argument. The correlations of the state $|\phi^+\rangle$ are the correlations of Einstein–Podolsky–Rosen bananas.

**Quantum Dynamics**

Quantum states change and evolve in time as systems interact dynamically, and the dynamics should preserve the Born probability rule for quantum states. The Born rule says that if you measure an observable with eigenvectors that include $|\alpha\rangle$ on
a quantum system in the state $|\psi\rangle$, the probability of finding the eigenstate $|\alpha\rangle$ is the square of the absolute value of the ‘length’ of the projection of $|\psi\rangle$ onto $|\alpha\rangle$, which is the square of the absolute value of the scalar product $\langle\alpha|\psi\rangle$. So when states change under a dynamical evolution defined by some operator $U(t)$ that transforms $|\psi\rangle \rightarrow U(t)|\psi\rangle$ and $|\alpha\rangle \rightarrow U(t)|\alpha\rangle$ in time $t$, the scalar product should be preserved, because the ‘length’ of a vector is defined by the scalar product.

If $U(t)$ operates on the state represented by the ket vector $|\alpha\rangle$, the corresponding operator on the bra vector $\langle\alpha|$ is the adjoint operator $U^*(t)$. You get $U^*(t)$ by transposing bras and kets in $U(t)$ and replacing $i$ by $-i$. So the bra vector of $U(t)|\alpha\rangle$ is $\langle\alpha|U^*(t)$ (the operator always operates on the flat side of a bra or ket).

The scalar product of $U(t)|\alpha\rangle$ and $U(t)|\psi\rangle$ is $\langle\alpha|U^*U|\psi\rangle$. The condition that the scalar product should be preserved under the transformation defined by $U(t)$ is then $\langle\alpha|U^*U|\psi\rangle = \langle\alpha|\psi\rangle$. This condition should hold for any states $|\alpha\rangle$ and $|\psi\rangle$, which means that $U^*(t)U(t) = I$, where $I$ is the identity operator. Equivalently, $U^{-1} = U^*$; the inverse of the operator defining the dynamics should be equal to the adjoint of the operator. Operators satisfying this condition are called unitary operators, so dynamical evolutions in quantum mechanics are unitary transformations.

Geometrically, this means that when two states transform under a dynamical evolution, the angle between the vectors remains the same, because it’s the angle that’s relevant for the probability according to the Born rule (the projection of one unit vector onto another depends only on the angle). Since orthogonality relations are preserved, a unitary transformation takes a set of orthogonal basis vectors to another set of orthogonal basis vectors.

To define a unitary operator, it’s sufficient to define the action of $U$ on a set of basis vectors. This fixes the transformation on any vector, because any vector can be represented in terms of its projections onto the basis vectors. For example, the 2-qubit transformation

$$
|0\rangle|0\rangle \rightarrow |0\rangle|0\rangle \\
|0\rangle|1\rangle \rightarrow |0\rangle|1\rangle \\
|1\rangle|0\rangle \rightarrow |1\rangle|1\rangle \\
|1\rangle|1\rangle \rightarrow |1\rangle|0\rangle
$$

is unitary, because the only change is that the second states in the product states $|1\rangle|0\rangle$ and $|1\rangle|1\rangle$, the states $|0\rangle$ and $|1\rangle$, are flipped. This is called a ‘controlled-not’ or CNOT transformation. The first qubit is the ‘control’ and the second qubit is the ‘target.’ The control qubit doesn’t change in the transformation. The target qubit doesn’t change if the control state is $|0\rangle$, but if the control state is $|1\rangle$, a ‘not’ operation is applied to the target, in the sense that $|0\rangle \rightarrow |1\rangle$ and $|1\rangle \rightarrow |0\rangle$. 

21
Figure S.1: Circuit diagrams showing the action of a CNOT gate. If the control is a superposition, as in the bottom diagram, the output is an entangled state.

A CNOT transformation can be part of a sequence of transformations in a quantum circuit. There’s a standard symbolism for representing various quantum transformations as ‘quantum gates’ in a circuit diagram. In a real circuit, these gates would be specific hardware devices. A CNOT gate is represented as in Figure S.1 by two circles, one solid, joined by a line. The diagrams are to be read from left to right, with input states on the left, and output states on the right. The top input ‘wire’ or channel on the left side of a gate carries the control qubit state, indicated by a subscript ‘c.’ The bottom input wire on the left side carries the target qubit state, indicated by a subscript ‘t.’ The two output wires or channels on the right side of a gate carry the output states. The way the gate functions is represented in the top two circuit diagrams for the control input states $|0\rangle_c$ and $|1\rangle_c$. The initial state of the target qubit is $|0\rangle_t$ in both cases here. The bottom circuit diagram shows how the gate functions when the input state of the control qubit is a linear superposition $|0\rangle_c + |1\rangle_c$. In that case, the output state is an entangled state of the control and target qubits on the output wires: $|0\rangle_c|0\rangle_t + |1\rangle_c|1\rangle_t$. That’s because the gate implements a unitary transformation, which is linear. So if $|0\rangle_c|0\rangle_t \to |0\rangle_c|0\rangle_t$ and $|1\rangle_c|0\rangle_t \to |1\rangle_c|1\rangle_t$, then $(|0\rangle_c + |1\rangle_c)|0\rangle_t \to |0\rangle_c|0\rangle_t + |1\rangle_c|1\rangle_t$. 
The first and third lines of the CNOT transformation

\[
\begin{align*}
|0\rangle_c |0\rangle_t &\xrightarrow{\text{CNOT}} |0\rangle_c |0\rangle_t \\
|1\rangle_c |0\rangle_t &\xrightarrow{\text{CNOT}} |1\rangle_c |1\rangle_t
\end{align*}
\]

provide a model for an idealized measurement interaction between two quantum systems. The measured qubit is the control qubit \(c\). The target qubit \(t\) represents the instrument ‘pointer.’ It starts out in the state \(|0\rangle_t\) (the pointer reads 0 or ‘ready’) and ends up mirroring the state of the control qubit (the pointer reads 0 or 1, depending on whether the state of the control is \(|0\rangle_c\) or \(|1\rangle_c\)). Now, because unitary transformations are linear, a CNOT transformation must transform a linear superposition or sum of \(|0\rangle_c |0\rangle_t\) and \(|1\rangle_c |0\rangle_t\) to the sum of the transformations for \(|0\rangle_c |0\rangle_t\) and \(|1\rangle_c |0\rangle_t\). So

\[
(|0\rangle_c + |1\rangle_c)|0\rangle_t = |0\rangle_c |0\rangle_t + |1\rangle_c |0\rangle_t \xrightarrow{\text{CNOT}} |0\rangle_c |0\rangle_t + |1\rangle_c |1\rangle_t
\]

This means that if the qubit being measured starts out in a linear superposition like \(|0\rangle + |1\rangle\), the final state of the combined system, measuring instrument plus measured system, ends up in an entangled state. You might reasonably expect that the final state should reflect what actually happens in a measurement, which is that with probability 1/2 the combined system ends up in the state \(|0\rangle|0\rangle\) and with probability 1/2 it ends up in the state \(|1\rangle|1\rangle\). The difference between what you get from the theory and what you expect is one way of putting the measurement problem of quantum mechanics. I take this up in section ??, The Measurement Problem, in the chapter Making Sense of It All.

The (not explicitly written) coefficients of the state of the measured qubit (the control in the CNOT transformation) are both \(1/\sqrt{2}\) here. But a qubit in any state with coefficients \(c_1, c_2\) could be measured by sending it through a CNOT gate:

\[
(c_1 |0\rangle + c_2 |1\rangle)|0\rangle \xrightarrow{\text{CNOT}} c_1 |0\rangle |0\rangle + c_2 |1\rangle |1\rangle
\]

The Pauli operators \(X, Y, Z\) in the section The Pauli Operators are self-adjoint and represent qubit observables. They are also unitary operators and so can represent dynamical transformations.

The Pauli operators are self-adjoint, \(X = X^*, Y = Y^*, Z = Z^*\), and unitarity is the condition \(XX^* = I, YY^* = I, ZZ^* = I\). So to show that the Pauli operators are unitary, all you need to prove is that \(X^2 = Y^2 = Z^2 = I\).

For \(X\), \((|0\rangle \langle 1| + |1\rangle \langle 0|) \cdot (|0\rangle \langle 1| + |1\rangle \langle 0|) = |0\rangle \langle 0| + |1\rangle \langle 1| = I\), because the two terms \(|0\rangle \langle 1| \cdot |0\rangle \langle 1|\) and \(|1\rangle \langle 0| \cdot |1\rangle \langle 0|\) are both zero, and \(|0\rangle \langle 1| \cdot |1\rangle \langle 0| = |0\rangle \langle 0|\) and \(|1\rangle \langle 0| \cdot |0\rangle \langle 1| = |1\rangle \langle 1|\). Similar calculations show that \(YY = I\) and \(ZZ = I\).
You can also see this by multiplying the matrix for each of these operators with itself. (See the section *States and Operators as Matrices*.)

Considered as a unitary transformation, the operator $X$ flips $|0\rangle$ and $|1\rangle$:  

\[
|0\rangle \xrightarrow{X} |1\rangle, |1\rangle \xrightarrow{X} |0\rangle 
\]

Explicitly: $(|0\rangle\langle 1| + |1\rangle\langle 0|) \cdot |0\rangle = 0 + |1\rangle = |1\rangle$ and $(|0\rangle\langle 1| + |1\rangle\langle 0|) \cdot |1\rangle = |0\rangle + 0 = |0\rangle$.

The operator $Z$ does nothing to $|0\rangle$ and multiples $|1\rangle$ by $-1$:  

\[
|0\rangle \xrightarrow{Z} |0\rangle, |1\rangle \xrightarrow{Z} -|1\rangle 
\]

That doesn’t change the states $|0\rangle$ and $|1\rangle$ as far as the probabilities defined by the Born rule are concerned, but it changes the relative phase of these two states in a general qubit state: $Z(\alpha|0\rangle + \beta|1\rangle) \rightarrow \alpha|0\rangle - \beta|1\rangle$. (See the section *Relative Phase*.)

The operator $Y$ implements the transformation  

\[
|0\rangle \xrightarrow{Y} i|1\rangle, |1\rangle \xrightarrow{Y} -i|0\rangle 
\]

You can think of the operator $iY$ as representing a transformation $X$ followed by a transformation $Z$, in that order, because $ZX = iY$ ($X$ followed by $Z$ operating on a state $\psi$ is $ZX|\psi\rangle$). To see this, multiply out the operator representatives of $Z$ and $X$ and show that this product is $iY$, or multiply the matrices of $Z$ and $X$ and show that the matrix product is equal to the matrix for $iY$. In fact, the Pauli operators are related by the conditions $XY = iZ$, $YZ = iX$, and $ZX = iY$ and $YX = -iZ$, $ZY = -iX$, and $XZ = -iY$.

For a photon, you can implement the transformations of the Pauli operators by passing the photon through an optical device that alters the polarization of the photon in an appropriate way. For an electron, $X$ rotates the spin of the electron by $180^\circ$ around the $x$-axis, $Y$ rotates the spin $180^\circ$ around the $Y$ axis, and $Z$ rotates the spin $180^\circ$ around the $z$-axis.

Although this plays no role in the book, it’s worth mentioning that the dynamics is implemented by a unitary operator that takes the form: $U = e^{iHt}$. The operator $H$ is a self-adjoint operator, called the Hamiltonian of the system. It represents the energy observable, with the possible energy values as eigenvalues. The Hamiltonian is unitary because the adjoint, $U^*$, is equal to $e^{-iH^*t}$, which is equal to $e^{-iHt}$, because $H$ is self-adjoint. Since $e^{iHt} \cdot e^{-iHt} = I$, $U^* = U^{-1}$, which is the definition of a unitary operator.

You can understand what this expression with an operator as exponent means via the expansion of $e^x$ as an infinite series: $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$, where $n!$
(‘$n$ factorial’) is defined as $1 \cdot 2 \cdot 3 \cdots n$. So $U = e^{iHt}$ is equivalent to an infinite sum of powers of operators: $I + iHt + \frac{(iHt)^2}{2!} + \frac{(iHt)^3}{3!} \cdots$. When applied to a quantum state, this gives you the change of the quantum state in time $t$. The rate of change at each instant is given by the differential equation:

$$i \frac{d|\psi\rangle}{dt} = H|\psi\rangle$$

This is Schrödinger’s equation.

---

**The bottom line**

- Quantum states change under dynamical evolutions defined by unitary operators. The unitary dynamics preserves the Born probability rule for quantum states.

- The Born rule says that if you measure an observable with eigenvectors that include $|\alpha\rangle$ on a quantum system in the state $|\psi\rangle$, the probability of finding the eigenstate $|\alpha\rangle$ is the square of the absolute value of the ‘length’ of the projection of $|\psi\rangle$ onto $|\alpha\rangle$. Geometrically, a unitary transformation ensures that when two states transform under a dynamical evolution, the angle between the vectors remains the same, because it’s the angle that’s relevant for the probability according to the Born rule (the projection of one unit vector onto another depends only on the angle). Since orthogonality relations are preserved, a unitary transformation takes a set of orthogonal basis vectors to another set of orthogonal basis vectors.

- An example of a unitary transformation is a controlled-not or CNOT transformation, which is a model for an idealized measurement interaction between two quantum systems.

- The Pauli operators $X, Y, Z$ are unitary operators and so can represent dynamical transformations as well as qubit observables. The operator $X$ flips $|0\rangle$ and $|1\rangle$. The operator $Z$ does nothing to $|0\rangle$ and multiplies $|1\rangle$ by $-1$. The operator $Y$ implements the transformation $|0\rangle \rightarrow i|1\rangle$ and $|1\rangle \rightarrow -i|0\rangle$. So $iY$ represents a transformation $X$ followed by a transformation $Z$. 

---

I’ve separated the following five topics—Relative Phase, Informationally Complete Observables, The Biorthogonal Decomposition Theorem, States and Opera-

25
tors as Matrices, and Mixed States—because they rather more challenging than the
previous sections. I do refer to them at various places in the book, so you might
need to dip into them as needed to follow the argument.

Relative Phase

I mentioned in the section *The Pauli Operators* that the global phase is irrelevant to
the state—$|\psi\rangle$ represents the same quantum state as $|\psi\rangle$ and generates the same
probabilities by the Born rule, and more generally $e^{i\theta}|\psi\rangle$ represents the same state
as $|\psi\rangle$. But the relative phase does make a difference. The vectors

$$|x\rangle = |0\rangle + |1\rangle$$

and

$$|\theta\rangle = |0\rangle + e^{i\theta}|1\rangle$$

represent different qubit states. An $X$-measurement on a qubit in the state $|x\rangle$
will produce the outcome $+1$ with probability 1, but an $X$-measurement on a qubit
in the state $|\theta\rangle$ will produce the outcome $+1$ with probability $\cos^2 \theta/2$ and $-1$ with
probability $\sin^2 \theta/2$.

It’s worth working through this example to see how probabilities are derived
from quantum states in the Dirac notation via the Born rule. First, express $|0\rangle$ and
$|1\rangle$ in the $\{|x\rangle, |x\rangle\}$ basis explicitly as:

$$|0\rangle = \frac{1}{\sqrt{2}}(|x\rangle + |x\rangle)$$

$$|1\rangle = \frac{1}{\sqrt{2}}(|x\rangle - |x\rangle)$$

Then

$$|\theta\rangle = \frac{1}{2}(|x\rangle + |x\rangle) + e^{i\theta}\frac{1}{2}(|x\rangle - |x\rangle)$$

$$= \frac{1 + e^{i\theta}}{2}|x\rangle + \frac{1 + e^{i\theta}}{2}|x\rangle$$

By the Born rule, the probability that a qubit in the state $|\theta\rangle$ subjected to an $X$-
measurement will produce the outcome 1 is:

$$\left| \frac{1 + e^{i\theta}}{2} \right|^2 = \left( \frac{1 + e^{i\theta}}{2} \right)\left( \frac{1 + e^{-i\theta}}{2} \right) = \frac{2 + e^{i\theta} + e^{-i\theta}}{4}$$

26
Since $e^{i\theta} + e^{-i\theta} = 2\cos\theta$, this becomes:

$$\frac{1 + 2\cos\theta}{2} = \cos^2\frac{\theta}{2}$$

The calculation for the probability of the outcome $-1$ is similar, or just observe that this probability is $1 - \cos^2\frac{\theta}{2} = \sin^2\frac{\theta}{2}$.

**The bottom line**
- The *global phase* is irrelevant to the state: $-|\psi\rangle$ represents the same quantum state as $|\psi\rangle$ and generates the same probabilities by the Born rule, and more generally $e^{i\theta}|\psi\rangle$ represents the same state as $|\psi\rangle$. But the *relative phase* does make a difference: $|\theta\rangle = |0\rangle + e^{i\theta}|1\rangle$ is not the same state as $|x+\rangle = |0\rangle + |1\rangle$.
- An explicit calculation in the Dirac notation shows how different probabilities are derived from the states $|\theta\rangle$ and $|x+\rangle$ via the Born rule.

**Informationally Complete Observables**

The expectation value or average value of an observable in a quantum state $|\psi\rangle$ is the weighted sum of the eigenvalues, with the probabilities defined by $|\psi\rangle$ as the weights. So the expectation value of $Z$, with eigenvalues $\pm 1$, in the state $|\psi\rangle$, denoted by $\langle Z \rangle_\psi$ is:

$$\langle Z \rangle_\psi = \text{prob}(1) - \text{prob}(-1) = |\langle 0 | \psi \rangle|^2 - |\langle 1 | \psi \rangle|^2$$

The probabilities are the absolute values of the ‘lengths’ of the projections of $|\psi\rangle$ onto the eigenstates $|0\rangle$ and $|1\rangle$ for the eigenvalues 1 and $-1$, according to the Born rule.

The expectation value can be derived from the representation of $Z$ as the operator $|0\rangle\langle 0| - |1\rangle\langle 1|$: 

$$\langle Z \rangle_\psi = \langle \psi | Z | \psi \rangle = \langle \psi | (|0\rangle\langle 0| - |1\rangle\langle 1|) | \psi \rangle = |\langle 0 | \psi \rangle|^2 - |\langle 1 | \psi \rangle|^2$$

In the tutorial for complex numbers in the section *Dirac’s Ingenious Idea*, I pointed out that the square of the absolute value of a complex number $c$ is $cc^*$, where $c^*$
is the complex conjugate of \( c \). Here \( \langle 0 | \psi \rangle = \langle \psi | 0 \rangle^* \) and \( \langle 1 | \psi \rangle = \langle \psi | 1 \rangle^* \), which explains the final step in the above derivation.

To see this, express the vector \(| \psi \rangle\) in terms of its components in the \(|0\rangle, |1\rangle\) basis as:

\[
| \psi \rangle = \langle 0 | \psi \rangle |0\rangle + \langle 1 | \psi \rangle |1\rangle
\]

and the corresponding bra vector as:

\[
\langle \psi | = \langle 0 | \psi \rangle^* \langle 0 | + \langle 1 | \psi \rangle^* \langle 1 |
\]

Taking the scalar product of \( \langle \psi | \) with \(|0\rangle\) on the right, you get:

\[
\langle \psi |0\rangle = \langle 0 | \psi \rangle^*
\]

because \( \langle 0 |0\rangle = 1 \) and \( \langle 1 |0\rangle = 0 \). Similarly, \( \langle \psi |1\rangle = \langle 1 | \psi \rangle^* \).

With an appropriate choice of the basis states \(|0\rangle\) and \(|1\rangle\) (so an appropriate choice of the observable \( Z \)), and ignoring the global phase, a general qubit can be represented as a unit vector:

\[
| \psi \rangle = \cos \theta |0\rangle + e^{i\phi} \sin \theta |1\rangle
\]

The projection operator onto this state is:

\[
| \psi \rangle \langle \psi | = (\cos \theta |0\rangle + e^{i\phi} \sin \theta |1\rangle)(\cos \theta \langle 0 | + e^{-i\phi} \sin \theta \langle 1 |)
\]

\[
= \cos^2 \theta |0\rangle \langle 0 | + \sin^2 \theta |1\rangle \langle 1 | + e^{i\phi} \cos \theta \sin \theta |0\rangle \langle 1 | + e^{-i\phi} \cos \theta \sin \theta |0\rangle \langle 1 |
\]

The three expectation values for the Pauli observables in the state \(| \psi \rangle\) turn out to be:

\[
\langle X \rangle_\psi = \cos \theta \sin \theta \cos \phi
\]

\[
\langle Y \rangle_\psi = \cos \theta \sin \theta \sin \phi
\]

\[
\langle Z \rangle_\psi = \cos^2 \theta - \sin^2 \theta
\]

and \( | \psi \rangle \langle \psi | \) can be expressed in terms of these expectation values as:

\[
\frac{1}{2}(I + \langle X \rangle_\psi X + \langle Y \rangle_\psi Y + \langle Z \rangle_\psi Z)
\]

where \( I \) is the identity operator: \( I = |0\rangle \langle 0 | + |1\rangle \langle 1 |\).

You can check this by showing that, in the expression for \( | \psi \rangle \langle \psi |\),

\[
\cos^2 \theta |0\rangle \langle 0 | + \sin^2 \theta |1\rangle \langle 1 | = \frac{1}{2}(I + \langle Z \rangle_\psi Z)
\]
So the projection operator onto the state of a qubit is specified by the expectation values of the three Pauli observables, $X, Y, Z$. This means that the quantum state of a qubit can be determined experimentally to arbitrary accuracy if you have a beam of qubits, all in the same quantum state $|\psi\rangle$. Take a sample of a sufficiently large number of qubits in the beam and divide the sample into three sets, and separately measure $X, Y,$ and $Z$ on the qubits in the sets. Then you can determine the expectation values $\langle X \rangle_\psi, \langle Y \rangle_\psi, \langle Z \rangle_\psi$ from these measurements to an accuracy that depends on the size of the sample. The expectation values are three real numbers that are sufficient to define the quantum state of the qubits in the beam.

It’s a general feature of quantum systems, not just qubits, that the state of a system can be specified by the probabilities of the outcomes of measurements on some ‘informationally complete’ or ‘fiducial’ set of observables. The probabilities (which determine the expectation values) can be estimated from the statistics of a sufficiently large number of measurements. Of course, in this process of ‘state tomography,’ as it’s called, you need a set of quantum systems all in the same quantum state. The number of observables in the informationally complete set will depend on the dimension of the state space. For a qubit, the three Pauli observables are an informationally complete set. For other quantum systems you need more than three observables. A set of informationally complete observables is not unique: any observable can be part of some informationally complete set.

The Biorthogonal Decomposition Theorem

There’s a really useful theorem about entangled states, the biorthogonal decomposition theorem or Schmidt decomposition theorem. The theorem says that for any pure state $|\Phi\rangle$ of a binary quantum system $S_1 + S_2$ (where $S_1$ and $S_2$ could each be composed of subsystems), there is an orthogonal set of unit vectors $|u_1\rangle, |u_2\rangle, \ldots$. 
associated with a basis in the state space of $S_1$, and a corresponding orthogonal set of unit vectors $|v_1\rangle, |v_2\rangle, \ldots$ for $S_2$ associated with a basis in the state space of $S_2$, such that $|\Phi\rangle$ can be expressed as a linear superposition of product states $|u_i\rangle|v_i\rangle$, for $i = 1, 2, \ldots$, so a ‘biorthogonal’ decomposition. The sum in the superposition could reduce to just one term, in which case $|\Phi\rangle$ is just a product state, but more generally the sum is an entangled state, a superposition of product states.

If the absolute values of the coefficients of the linear superposition are all different, the biorthogonal decomposition is unique. Otherwise, the biorthogonal decomposition is not unique. (If a coefficient is a real number $\pm c$, the absolute value is $c$. If the coefficient is a complex number, the absolute value is $\sqrt{cc^*}$.)

For example, the entangled two-qubit Bell state $|\phi^+\rangle = |0\rangle|0\rangle + |1\rangle|1\rangle$ is a biorthogonal decomposition of the state $|\phi^+\rangle$ into products of corresponding computational basis states for the two qubits. In this case, the absolute values of the coefficients are the same, so the biorthogonal decomposition is not unique. The state $|\phi^+\rangle$ could equally well be expressed as a biorthogonal decomposition of $X$-eigenstates of the two qubits $|\phi^+\rangle = |+\rangle|+\rangle + |-\rangle|-\rangle$ or, in fact, as a biorthogonal decomposition with respect to any orthogonal basis states, $|\phi^+\rangle = |s_1\rangle|s_1\rangle + |s_2\rangle|s_2\rangle$, where the basis $|s_1\rangle, |s_2\rangle$ is related to the computational basis by a unitary transformation with real coefficients.

That’s the case if you transform from the computational basis to the $X$-basis where the coefficients are real numbers. If you transform to a basis like the $Y$-basis that is related to the computational basis by a unitary transformation with complex coefficients, $|y_+\rangle = |0\rangle + i|1\rangle, |y_-\rangle = |0\rangle - i|1\rangle$, then you get $|y_+\rangle|y_-\rangle + |y_-\rangle|y_+\rangle$, which is the Bell state $|\psi^-\rangle$ rather than $|\phi^+\rangle$.

To see why the theorem might be useful, suppose Alice and Bob share two qutrits, and consider two bases, $|1\rangle, |2\rangle, |3\rangle$ and $|1\rangle', |2\rangle', |3\rangle'$, where $|1\rangle = |1\rangle'$ and $|2\rangle', |3\rangle'$ are rotated relative to $|2\rangle, |3\rangle$ in the plane orthogonal $|1\rangle$. See Figure S.2.

![Figure S.2: Two bases with a common state vector](image-url)
The maximally entangled state
\[ |\Phi\rangle = |1\rangle_A |1\rangle_B + |2\rangle_A |2\rangle_B + |3\rangle_A |3\rangle_B \]
has equal coefficients for all the terms in the biorthogonal decomposition. If Alice
and Bob both measure the observable \( O \) with eigenstates \( |1\rangle, |2\rangle, |3\rangle \) of \( O \), then
they will both get the same outcome, either the outcome corresponding to \( |1\rangle \), or
the outcome corresponding to \( |2\rangle \), or the outcome corresponding to \( |3\rangle \), with equal
probability 1/3. (As usual, I’ve left out the equal coefficient \( 1/\sqrt{3} \) for each of
the terms in the state \( |\Phi\rangle \).) Because the coefficients are all equal, the biorthogonal
decomposition of \( |\Phi\rangle \) is not unique, and \( |\Phi\rangle \) could equally well be expressed as
\[ |\Phi\rangle = |1\rangle_A |1\rangle_B + |2\rangle_A |2\rangle_B + |3\rangle_A |3\rangle_B = |1\rangle_A |1\rangle_B + |2\rangle_A |2\rangle_B + |3\rangle_A |3\rangle_B \]
So if Alice and Bob both measure the observable \( O' \) with eigenstates \( |1\rangle' = |1\rangle, |2\rangle', |3\rangle' \), they will again get the same outcome with equal probability.

What if Alice measures \( O \) and Bob measures \( O' \)? Then Alice will get the
outcome corresponding to \( |1\rangle \) for \( O \) if and only if Bob gets the corresponding
outcome for \( O' \). If Alice gets one of the other outcomes, then Bob will too, but
not necessarily the outcome corresponding to Alice’s outcome (so Alice could get
the outcome corresponding to the eigenvector \( |2\rangle \) and Bob could get the outcome
corresponding to \( |3\rangle' \), or Alice could get the outcome corresponding to \( |3\rangle \) and Bob
could get the outcome corresponding to \( |2\rangle' \)).

To see why, express Bob’s states in the biorthogonal decomposition as \( O'\)-
eigenstates. Suppose
\[
|1\rangle = |1\rangle' \\
|2\rangle = a|2\rangle' + b|3\rangle' \\
|3\rangle = c|2\rangle' + d|3\rangle'
\]
where all the coefficients \( a, b, c, d \) are real numbers. Then
\[
|\Phi\rangle = |1\rangle_A |1\rangle_B + |2\rangle_A |2\rangle_B + |3\rangle_A |3\rangle_B \\
= |1\rangle_A |1\rangle_B + |2\rangle_A (a|2\rangle' + b|3\rangle') + |3\rangle_A (c|2\rangle' + d|3\rangle') \\
= |1\rangle |1\rangle' + a|2\rangle |2\rangle' + b|2\rangle |3\rangle' + c|3\rangle |2\rangle' + d|3\rangle |3\rangle'
\]
Remember that all these terms should be multiplied by the coefficient \( 1/\sqrt{3} \).
According the Born rule, then, the probability that Alice gets the outcome corre-
sponding to \( |1\rangle \) for \( O \) and Bob gets the corresponding outcome \( |1\rangle' = |1\rangle \) for \( O' \) is
1/3. There’s also a probability \( |a|^2/3 \) that Alice gets the outcome corresponding to
\( |2\rangle \) and Bob gets the outcome corresponding to \( |2\rangle' \), a probability \( |b|^2/3 \) that Alice
gets the outcome corresponding to $|2\rangle'$ and Bob gets the outcome corresponding to $|3\rangle'$, a probability $|c|^2/3$ that Alice gets the outcome corresponding to $|3\rangle$ and Bob gets the outcome corresponding to $|2\rangle'$, and a probability $|d|^2/3$ that Alice gets the outcome corresponding to $|3\rangle$ and Bob gets the outcome corresponding to $|3\rangle'$. In other words, if Alice and Bob measure two different observables that have a common eigenvector, then either they will both get the outcome corresponding to the common eigenvector, or they will both get some other outcome. The probability is zero that Alice will get the outcome corresponding to the common eigenvector and Bob won’t, or that Bob will and Alice won’t.

Now consider an analogous but slightly more complicated case that comes up in the quantum simulation of Aravind–Mermin bananas in the More section of Chapter ??, Quantum Magic. Alice and Bob share three pairs of qubits, each in the entangled state $|\phi^+\rangle = |0\rangle|0\rangle + |1\rangle|1\rangle$, so the six-qubit state is

$$|\phi^+\rangle \otimes |\phi^+\rangle \otimes |\phi^+\rangle = \frac{1}{\sqrt{8}} (|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B) \otimes (|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B) \otimes (|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B)$$

where I’ve indicated the Alice-states and Bob-states by subscripts.

If you multiply out the terms in this expression, you get a linear combination of eight terms like $|000\rangle_A |000\rangle_B$ (this is the product of the first three terms in the three states $|\phi^+\rangle$). Taking all the Alice-states together and all the Bob states together, these terms can be re-ordered as products of Alice-states and Bob-states, like

$$|000\rangle_A |000\rangle_B$$

where $|000\rangle_A$ is short for $|0\rangle_A |0\rangle_A |0\rangle_A$. The whole expression is

$$|\phi^+\rangle |\phi^+\rangle |\phi^+\rangle = |000\rangle_A |000\rangle_B + |001\rangle_A |001\rangle_B + \cdots$$

for all of the eight triples of bits 000, 001, 010, 011, 100, 101, 110, 111.

This is a biorthogonal decomposition for the Alice and Bob systems (which each consists of three qubits) with equal coefficients in the $ZZZ$ basis, so the state will take the same form in the $XXX$ basis. Now suppose Alice measures her three qubits in the $Z$-basis, so she measures $ZZZ$, and Bob measures his first qubit in the $Z$-basis but his other two qubits in the $X$-basis, so he measures $ZXX$. (This is the sort of observable relevant to the simulation of Aravind–Mermin bananas.) As in the previous case, the biorthogonal decomposition ensures that Alice and Bob will get the same outcomes for the $Z$-measurements on their first qubits.
To work out the probabilities explicitly, you would express Alice’s and Bob’s three-qubit states in terms of $Z$-eigenstates and $X$-eigenstates, as appropriate. So

$$|\phi^+\rangle|\phi^+\rangle|\phi^+\rangle = |000\rangle_A|0\rangle_B(|+\rangle_B + |\rangle_B)(|+\rangle_B + |\rangle_B) + \cdots$$

In the expression after the equal sign, I’ve replaced each of Bob’s second and third qubit states $|0\rangle_B$ in the first term by $|+\rangle_B + |\rangle_B$. The $\cdots$ refers to the other seven product states. For the next state $|001\rangle_A|001\rangle_B$, Bob’s second qubit state $|0\rangle_B$ is replaced by $|+\rangle_B + |\rangle_B$, and his third qubit state $|1\rangle_B$ is replaced by $|+\rangle_B - |1\rangle_B$, and so on.

Now what’s relevant here is that you get four states in the sum of the form $|0\cdots\rangle_A|0\cdots\rangle_B$ and four states of the form $|1\cdots\rangle_A|1\cdots\rangle_B$. Here $|0\cdots\rangle_A$ are the four states $|000\rangle_A$, $|001\rangle_A$, $|010\rangle_A$, and $|011\rangle_A$ and $|1\cdots\rangle_A$ are the four states $|100\rangle_A$, $|101\rangle_A$, $|110\rangle_A$, and $|111\rangle_A$. The states of the form $|\cdots\rangle_B$ are either $|+\rangle_B + |\rangle_B$ or $|+\rangle_B - |\rangle_B$. So, again, it follows from the Born rule for quantum probabilities that Alice gets the outcome corresponding to the state $|0\rangle$ for the $Z$-measurement on her first qubit if and only if Bob gets the corresponding outcome for the $Z$-measurement on his first qubit, and she gets the outcome corresponding to the state $|1\rangle$ on her first qubit if and only if Bob gets the corresponding outcome for the $Z$-measurement on his first qubit, and each of these possibilities occurs with probability $1/2$.

**The bottom line**

- The biorthogonal decomposition theorem says that you can always express the state a composite system $AB$, as a linear superposition of product states for $A$ and $B$, where the $A$-states and the $B$ states are orthogonal states associated with a basis in the $A$ state space and a basis in the $B$ state space, respectively.

- For example, the entangled 2-qubit Bell state $|\phi^+\rangle = |0\rangle|0\rangle + |1\rangle|1\rangle$ is a biorthogonal decomposition.

- The theorem has many applications. In particular, it gets applied in the quantum simulation of Aravind–Mermin bananas in subsection ??, Simulating Aravind–Mermin Bananas, in the More section of the Quantum Magic chapter. I work out an example relevant to the simulation of Aravind–Mermin bananas explicitly.
States and Operators as Matrices

Operators can be represented by matrices. I don’t explicitly use matrices in this book, but it’s useful to see how Dirac’s ‘bra-ket’ notation for an operator is related to the rows and columns of the corresponding matrix. In particular, it’s useful to see this for the Pauli operators.

The idea behind matrices is pretty straightforward. A pair of linear equations like

\[
\begin{align*}
x' &= ax + by \\
y' &= cx + dy
\end{align*}
\]

can be expressed as a matrix equation, in which the single column matrix \(\begin{pmatrix} x \\ y \end{pmatrix}\) is operated on, or multiplied by, a four-element square matrix to produce the new single column matrix \(\begin{pmatrix} x' \\ y' \end{pmatrix}\):

\[
\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}
\]

The matrix multiplication rule simply follows the multiplication and addition operations in the linear equations. To get the first element of the new single column matrix, \(x'\), multiply the two elements in the first row of the four-element square matrix by the two elements in the single column matrix sequentially and add the products. To get the second element, \(y'\), multiply the two elements in the second row of the four-element matrix by the two elements in the single column matrix sequentially and add the products.

To multiply two four-element square matrices follow the same rule for each element in the resulting product matrix. The element in the first row of the first column is obtained by multiplying the two elements in the first row of the first matrix with the two elements in the first column of the second matrix sequentially and adding the products. The element in the first row of the second column is obtained by multiplying the two elements in the first row of the first matrix with the two elements in the second column of the second matrix sequentially and adding the products. So

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}
\]

In matrix notation, states represented by ket vectors in a qubit state space are represented by single column matrices (so they are referred to as column vectors), bra vectors by single row matrices (so row vectors), and operators by four-element
square matrices. The computational basis ket vectors $|0\rangle$ and $|1\rangle$ are represented by the column vectors

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and the corresponding bra vectors are represented as the row vectors

$$\langle 0| = (1, 0), \quad \langle 1| = (0, 1)$$

The matrix associated with the operator $|0\rangle\langle 1|$ is a four-element square matrix obtained as the product

$$|0\rangle\langle 1| = (1, 0) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

To get the first column of the four-element matrix, multiply the first element 1 in the row matrix with the column matrix $(0 \ 1)$. To get the second column of the 4-element matrix, multiply the second element 0 in the row matrix with the column matrix $(0 \ 1)$. Similarly,

$$|1\rangle\langle 0| = (0, 1) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

So the Pauli operator $X = |0\rangle\langle 1| + |1\rangle\langle 0|$ can be expressed in matrix form as

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

and the Pauli operators $Y$ and $Z$ can be expressed as

$$Y = -i|0\rangle\langle 1| + i|1\rangle\langle 0| = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

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

The identity operator $I$ becomes

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

Here’s another way of thinking about this. The top and bottom row of a four-element matrix can be numbered as the 0 row and the 1 row, and the left and right columns as the 0 column and the 1 column. The two elements in the top row are
the 00 element (row 0, column 0) and the 01 element (row 0, column 1). The
two elements in the bottom row are the 10 element (row 1, column 0) and the 11
element (row 1, column 1). The matrix elements of an operator depend on the basis
chosen. In the standard or computational basis, the matrix elements of a general
qubit operator $Q$ on the top row are $\langle 0 | Q | 0 \rangle$ for the 00 element and $\langle 0 | Q | 1 \rangle$ for
the 01 element, and the matrix elements on the bottom row are $\langle 1 | Q | 0 \rangle$ for the
10 element and $\langle 1 | Q | 1 \rangle$ for the 11 element. So for the operator $|0\rangle\langle 1|$, the matrix
elements in the top row are $\langle 0 | 0 \rangle \langle 1 | 0 \rangle = \langle 1 | 0 \rangle = 0$ and $\langle 0 | 0 \rangle \langle 1 | 1 \rangle = \langle 1 | 1 \rangle = 1$,
and the matrix elements in the bottom row are $\langle 1 | 0 \rangle \langle 1 | 0 \rangle = 0$ and $\langle 1 | 0 \rangle \langle 1 | 1 \rangle = 0$.

A tensor product state of two qubits is represented by a four-element column
vector

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

You multiply the top element 1 in the first column vector by the elements in the
second column vector to get the first two rows in the four-element column vector
(the first and second elements, reading down), and you multiply the bottom element
0 in the first column vector by the elements in the second column vector to get the
second two rows in the 4-element column vector (the third and fourth elements,
reading down).

Similarly,

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

so

$$|\psi^-\rangle = |0\rangle |1\rangle - |1\rangle |0\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$$
Operators can be represented by matrices. Although matrices are not used in this book, this section is useful in showing how Dirac’s ‘bra-ket’ notation for an operator is related to the rows and columns of the corresponding matrix.

In particular, it’s useful to see this for the Pauli operators, and for the identity operator.

Mixed States

A qubit can be in a pure state that is a linear superposition of other pure states, like the pure state $|+\rangle = |0\rangle + |1\rangle$. It can also be in a mixture of pure states. For example, a qubit could be prepared in such a way that it is either in the pure state $|0\rangle$ with probability $p_0$ or in the pure state $|1\rangle$ with probability $p_1$. You could have a whole collection of qubits prepared in this way, and sampling a qubit from such a mixture of qubit states would produce a qubit in the state $|0\rangle$ with probability $p_0$ or a qubit in the state $|1\rangle$ with probability $p_1$. The probabilities $p_1$ and $p_2$ are referred to as the ‘weights’ of the pure states in the mixture.

The expectation values or average values of an observable $Q$ in the pure states $|0\rangle$ and $|1\rangle$ are $\langle Q \rangle_{|0\rangle} = \langle 0|Q|0\rangle$ and $\langle Q \rangle_{|1\rangle} = \langle 1|Q|1\rangle$. So the expectation value of $Q$ in a mixture of states $|0\rangle$ and $|1\rangle$ with weights $p_1$ and $p_2$ is the weighted sum of the expectation values in the two states: $\langle Q \rangle = p_0 \langle 0|Q|0\rangle + p_1 \langle 1|Q|1\rangle$.

There’s a neat way of expressing the expectation value of an observable $Q$ in a mixture in terms of an operator called the ‘density operator’ of the mixture, and a function called the ‘trace’ function, represented by $\text{Tr}(\ldots)$:

$$\langle Q \rangle_{\rho} = \text{Tr}(\rho Q)$$

where $\rho = p_0 |0\rangle \langle 0| + p_1 |1\rangle \langle 1|$ is the ‘density operator’ representing the mixture. The density operator of a mixture of orthogonal pure states is the weighted sum of the projection operators onto the pure states in the mixture, where the weights are the weights of the pure states in the mixture. (I’ve been representing operators by capital Roman letters, but it’s usual to represent a density operator by the symbol $\rho$, which is also used to represent a probability distribution or mixture of classical pure states.)

Here’s how to see this. The trace of an operator $R$ is the sum of the elements along the diagonal of the matrix representing $R$. For a qubit, a matrix representing an operator has four elements, and these elements depend on the basis. The two
diagonal elements are $\langle 0| R |0 \rangle$ and $\langle 1| R |1 \rangle$ in the computational basis. So the trace of $R$ is $\text{Tr}(R) = \langle 0| R |0 \rangle + \langle 1| R |1 \rangle$.

The trace is an invariant function of an operator: it takes the same value for any basis. For example, with respect to the basis $|+, -\rangle$, $\text{Tr}(R) = \langle +| R |+ \rangle + \langle -| R |- \rangle$. Now, $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. So the first term, $\langle +| R |+ \rangle$, is the same, except that the off-diagonal terms $|0\rangle |1\rangle$ and $|1\rangle |0\rangle$ appear with a $-i$ sign and so cancel out with the corresponding elements in $\langle +| R |+ \rangle$.

The trace is a linear function: $\text{Tr}(cR) = c \text{Tr}(R)$, and $\text{Tr}(R + S) = \text{Tr}(R) + \text{Tr}(S)$. So the trace of $\rho Q = \langle p_0 |0\rangle \langle 0| + p_1 |1\rangle \langle 1| Q$ is $p_0 \text{Tr}(\langle 0| Q |0\rangle) + p_1 \text{Tr}(\langle 1| Q |1\rangle)$. Since the trace is invariant with respect to the choice of basis, you can choose any convenient basis to calculate the trace. Taking the trace with respect to the computational basis, the trace of $|0\rangle \langle 0| Q$ is $\langle 0|0\rangle \langle 0| Q |0\rangle + \langle 1|0\rangle \langle 0| Q |1\rangle = \langle 0| Q |0\rangle$ because $\langle 0|0\rangle = 1$ and $\langle 1|0\rangle = 0$. Similarly, the trace of $|1\rangle \langle 1| Q$ is $\langle 1| Q |1\rangle$. So $\text{Tr}(\rho Q) = p_0 \langle 0| Q |0\rangle + p_1 \langle 1| Q |1\rangle$, which is the expectation value of $Q$ in the mixture.

From the density operator for a mixture, you can also calculate the probabilities of the eigenvalues or possible values of an observable $Q$. Suppose $Q$ is a qubit observable with two possible values, $q_1$ and $q_2$. Then $Q$ can be expressed in the spectral representation as $Q = q_1 P_{q_1} + q_2 P_{q_2}$, where the operators $P_{q_1}$ and $P_{q_2}$ are the projection operators onto the eigenvectors corresponding to the eigenvalues $q_1$ and $q_2$. (See the section The Pauli Operators.) In terms of the density operator $\rho$, you can express the expectation value of $Q$ as $\text{Tr}(\rho Q) = \text{Tr}(\rho (q_1 P_{q_1} + q_2 P_{q_2})) = q_1 \text{Tr}(\rho P_{q_1}) + q_2 \text{Tr}(\rho P_{q_2})$. This is the weighted sum of the eigenvalues of $Q$ with the weights $\text{Tr}(\rho P_{q_1})$ and $\text{Tr}(\rho P_{q_2})$, so these weights are the probabilities of the eigenvalues.

Since you can derive the probabilities and expectation values of any observable in a mixture from the density operator, this operator plays the same role for mixtures as the state vector does for pure states. A quantum system in a mixture is said to be in a mixed state represented by density operator of the mixture, and the Born rule for a mixed state $\rho$ is

$$\text{prob}_\rho(q) = \text{Tr}(\rho P_q)$$

$$\langle Q \rangle_\rho = \text{Tr}(\rho Q)$$

The Born rule in this form is quite general and applies to pure and mixed states. For a pure state $|\psi\rangle$, the density operator $\rho$ representing the state is the projection operator $P_\psi = |\psi\rangle \langle \psi|$ onto the state. Then $\text{prob}_\psi(q) = \text{Tr}(\langle \psi | q \rangle \langle q | \psi \rangle) = \text{Tr}(|\psi\rangle \langle \psi | q \rangle \langle q | \psi \rangle)$. To calculate the trace, use the basis defined by the eigenvectors of $Q$. Then $\text{Tr}(\langle \psi | q \rangle \langle q | \psi \rangle)$ = $\langle q | \psi \rangle \langle \psi | q \rangle + \text{terms that are zero} = \text{prob}_\psi(q)$.
\langle q|\psi\rangle\langle\psi|q\rangle = |\langle q|\psi\rangle|^2$. This is the probability of $q$ in the state $|\psi\rangle$ according to the Born rule: the square of the absolute value of the projection of $|\psi\rangle$ onto $|q\rangle$. The terms that are zero are terms like $\langle q'|\psi\rangle\langle\psi|q\rangle\langle q|q'\rangle$ for eigenvectors orthogonal to $|q\rangle$, which are zero because $\langle q|q'\rangle = 0$.

For the expectation value $\langle Q \rangle_\psi = \text{Tr}(|\psi\rangle\langle\psi|Q)$, take the trace with respect to any basis of orthogonal unit vectors containing $|\psi\rangle$. Then $\text{Tr}(|\psi\rangle\langle\psi|Q|\psi\rangle = \langle\psi|Q|\psi\rangle$, because $|\psi\rangle$ is a unit vector and so $\langle\psi|\psi\rangle = 1$.

Different mixtures can be represented by the same mixed state density operator. For example, the mixture of pure states $|0\rangle, |1\rangle$ with equal weights is represented by the same mixed state density operator as the mixture of pure states $|+\rangle, |−\rangle$ with equal weights

$$\rho = \frac{1}{2} |0\rangle\langle 0| + \frac{1}{2} |1\rangle\langle 1| = \frac{1}{2} |+\rangle\langle +| + \frac{1}{2} |−\rangle\langle −|$$

You can see this by writing out the ket and bra vectors $|+\rangle, \langle +|$ and $|−\rangle, \langle −|$ explicitly in terms of the ket and bra vectors $|0\rangle, \langle 0|$ and $|1\rangle, \langle 1|$ and multiplying out the expressions

$$\frac{1}{2}(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)) \cdot \frac{1}{\sqrt{2}}(|0\rangle + \langle 1|) + \frac{1}{2}(\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)) \cdot \frac{1}{\sqrt{2}}(|0\rangle - \langle 1|)$$

The $|0\rangle\langle 1|$ and $|1\rangle\langle 0|$ terms cancel out because they appear with $+$ signs in the expression you get from the first product and with $−$ signs in the expression you get from the second product. The remaining terms add up to

$$\frac{1}{2}(\frac{1}{\sqrt{2}}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|) + \frac{1}{2}(\frac{1}{\sqrt{2}}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|) = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$$

Since the identity operator $I$ can be expressed as $|0\rangle\langle 0| + |1\rangle\langle 1|$ or as $|+\rangle\langle +| + |−\rangle\langle −|$, the mixed state of these two mixtures is represented by the density operator $\rho = I/2$. In fact, any completely random mixture of orthogonal qubit pure states is represented by the mixed state density operator $I/2$, which also represents mixtures of nonorthogonal qubit pure states.

Measurement probabilities for a mixture of states are given by the Born rule via the density operator $\rho$ and don’t depend on the representation of $\rho$ as a specific mixture of pure states. So the fact that different mixtures of pure states are represented by the same density operator means that such mixtures are indistinguishable by any measurements. This is also a feature of mixtures with non-equal weights. You could also mix nonorthogonal pure states, and the indistinguishability will apply to these states as well.

This might seem odd because you could prepare an equal weight mixture of linearly polarized photons in states $|0\rangle, |1\rangle$ by sending photons through a beamsplitter with its axis of polarization in a certain direction $z$, or you could prepare an
equal weight mixture of diagonally polarized photons in states $|+\rangle$, $|−\rangle$ by passing photons through a beamsplitter with its axis of polarization rotated $45^\circ$ to the $z$ direction, and these two mixtures would seem to be quite different.

But consider what happens if you select a photon at random from the mixture represented by the mixed state $\frac{1}{2}|+\rangle\langle+| + \frac{1}{2}|−\rangle\langle−|$ and send it through a beamsplitter with its axis of polarization in the $z$ direction. If you select a photon at random from the mixture, it will either be in the state $|+\rangle$ with probability $1/2$ or in the state $|−\rangle$ with probability $1/2$. If a photon in the state $|+\rangle$ goes through the beamsplitter, it will end up in the state $|0\rangle$ with probability $1/2$ or in the state $|1\rangle$ with probability $1/2$, and similarly for the state $|−\rangle$. So the probability that a photon selected at random from the mixture and sent through the beamsplitter ends up in the state $|0\rangle$ is $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, and similarly the probability that the photon ends up in the state $|1\rangle$ is $\frac{1}{2}$. That’s the same as the probabilities for the mixture $\frac{1}{2}|0\rangle\langle0| + \frac{1}{2}|1\rangle\langle1|$. The same thing will happen for any polarization measurements, so the two mixtures are really indistinguishable.

Finally, suppose you have an entangled state of two qubits, $A$ and $B$, like

$$|\psi\rangle = c_1|0\rangle_A|0\rangle_B + c_2|1\rangle_A|1\rangle_B$$

where $c_1$ and $c_2$ are real number coefficients whose squares, representing probabilities, add up to 1. What’s the quantum state of qubit $B$ alone?

It should be a state $\rho_B$ such that

$$\langle \psi | I_A Q_B | \psi \rangle = \text{Tr}(\rho_B Q_B)$$

where $I_A$ is the identity operator for qubit $A$ and $Q_B$ is any observable of qubit $B$ alone. This expression is $I_A Q_B$ sandwiched between $\langle \psi | = c_1 A |0\rangle_B\langle0| + c_2 A |1\rangle_B\langle1|$ and $|\psi\rangle = c_1|0\rangle_A|0\rangle_B + c_2|1\rangle_A|1\rangle_B$. (We’ve written bra vectors as $A |0\rangle_B\langle0|$ and ket vectors as $|0\rangle_A, |0\rangle_B$.)

If you work this out, you get

$$c_1^2 A |0\rangle_B\langle0| I_A Q_B |0\rangle_A|0\rangle_B + c_2^2 A |1\rangle_B\langle1| I_A Q_B |1\rangle_A|1\rangle_B$$

$$+ c_1 c_2 (A |0\rangle_B\langle0| I_A Q_B |1\rangle_A|1\rangle_B + A |1\rangle_B\langle1| I_A Q_B |0\rangle_A|0\rangle_B)$$

The two terms with coefficients $c_1 c_2$ are

$$A |0\rangle_B\langle0| I_A Q_B |1\rangle_A|1\rangle_B = A |0\rangle_B\langle0| A^\dagger B |0\rangle_Q B |1\rangle_B = A |0\rangle_B\langle0| Q_B |1\rangle_B = 0$$

and

$$A |1\rangle_B\langle1| I_A Q_B |0\rangle_A|0\rangle_B = A |1\rangle_B\langle1| A^\dagger B |1\rangle_Q B |0\rangle_B = A |1\rangle_B\langle1| Q_B |0\rangle_B = 0$$
Both these terms are zero because $|0\rangle$ and $|1\rangle$ are orthogonal, so $A\langle 0|1 \rangle_A = A\langle 1|0 \rangle_A = 0$.

The other two terms are equal to

$$c_1^2 A\langle 0|I_A|0 \rangle_A \cdot B\langle 0|Q_B|0 \rangle_B + c_2^2 A\langle 1|I_A|1 \rangle_A \cdot B\langle 1|Q_B|1 \rangle_B$$

$$= c_1^2 A\langle 0|0 \rangle_A \cdot B\langle 0|Q_B|0 \rangle_B + c_2^2 A\langle 1|1 \rangle_A \cdot B\langle 1|Q_B|1 \rangle_B$$

and this is just

$$c_1^2 B\langle 0|Q_B|0 \rangle_B + c_2^2 B\langle 1|Q_B|1 \rangle_B$$

because $A\langle 0|0 \rangle_A = A\langle 1|1 \rangle_A = 1$. So

$$\rho_B = c_1^2 |0\rangle\langle 0| + c_2^2 |1\rangle\langle 1|$$

where I’ve left out the subscripts $B$.

This is a mixed state, called the ‘reduced state’ of the subsystem $B$, representing a mixture of the states $|0\rangle$ and $|1\rangle$ with weights $p_1 = c_1^2$ and $p_2 = c_2^2$. Even though the two-qubit composite state is an entangled pure state, the state of a subsystem is a mixed state. This will always be the case, for an entangled pure state of any composite system, not necessarily two qubits, unless the state of the composite system is a product state.
The bottom line

- A quantum system can be in a pure state that is a linear superposition of other pure states, or you can have a whole collection of quantum systems, called a mixture, prepared in different states that occur in the mixture with different probabilities or ‘weights.’

- There’s a neat way of expressing the expectation value of an observable in a mixture, or the probabilities of the possible values of an observable, in terms of an operator called the ‘density operator’ of the mixture.

- The density operator plays the same role for mixtures as the state vector does for pure states. A quantum system selected at random from a mixture is said to be in a mixed state represented by density operator of the mixture.

- Different mixtures can be represented by the same mixed state density operator, and such mixtures are indistinguishable by any measurements.

- If you have two (or more) quantum systems in an entangled pure state, the state of a subsystem is in general a mixed state, called the ‘reduced state’ of the subsystem.