Quantum Information

Kasper Peeters (based on notes by Douglas J Smith and Daniele Dorigoni)

February 17, 2024

A mobile-friendly version of these notes is available at
https://www.maths.dur.ac.uk/users/kasper.peeters/qcomp/2023/

Lecture notes for the 2023-2024 module at Durham University.

Contents

1

Introduction

LECTURE 1

1.1 A bit of history

While it perfectly possible to study quantum information and quantum computing as isolated mathematical topics without any reference their origin in physics, this would ignore a large part of interesting historical development, and it would also fail to catch the actual real-world importance it may one day have. So before we go into the underlying maths, this chapter will put things into context and try to give some idea about what you can and cannot do with quantum information and quantum computing.

Historically, the field is now almost 45 years old. The first suggestion of using a quantum version of the universal Turing machine goes back to the work of Paul Benioff [1] in 1980. Richard Feynman [3] and Yuri Manin then independently came to the conclusion that simulating the quantum world on a classical computer is very limiting (for reasons we will explore later) and those limitations might be avoidable by using a computer based on quantum mechanics itself instead. Theoretical work in that direction, namely the development of *algorithms* that could be run on such – as of then non-existent – quantum computers where developed by many people starting with the work of David Deutsch [2] and Peter Shor in the early 1980's. It then took some four decades to get to actual physical realisations of quantum computers, and even those are still quite limited in size and capability.

But before we go deeper into the details of those developments, and get lost in mathematical details that sometimes hide the things that are really important, it is good to think briefly about *why* quantum computing might be different from classical computing. In physics terms, there are two key things in quantum mechanics which make it different from classical mechanics: *superposition* and *entanglement*. These two things simply do not exist in classical systems. We will discuss these in some more detail below to refresh your memory.

Most of the notes here will discuss what is known as "quantum information": how you store and manipulate information in systems which are built from quantum-mechanical ingredients. The second part of the module will then use this to study quantum computing: how can we do computations by manipulating quantum information, and how do those computations differ from classical computations.

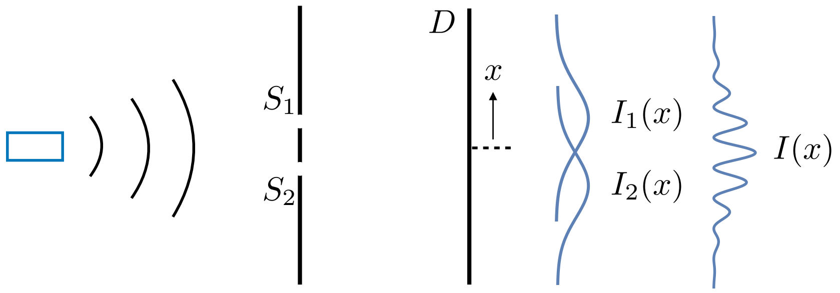


Figure 1.1: The double slit experiment, in which the quantum particle goes through two slits *at the same time*.

1.2 Quantum is different

1.2.1 Superposition: doing multiple things at once

A typical course on quantum mechanics introduces the need for something beyond the classical world by discussing the double-slit experiment. This famous experiment involves a single source emitting individual particles or photons. Their path is blocked by a screen with two slits, and behind that screen sits a detector. Even though we can see particle-like behaviour of the quanta because individual blobs appear on the detector one-by-one, the pattern that emerges after a large number of quanta have been emitted is one of interference.

Superposition means that quantum systems "compute" multiple classical things at the same time.

What this shows is that quantum particles, in a way, go through both slits *at once*, and only the measurement at the detector forces them to become classical again, giving a concrete outcome for the position. The quantum world thus makes the particle do two things *at the same time*, only collapsing onto a definite prediction once the measurement is made. The state of the system, until the time that the measurement is made, is one of a superposition of the particle going through slit S_1 and another one of the particle going through slit S_2.

Application of superposition: quantum algorithms and quantum computing.

It is this "doing multiple things at once" aspect that underlies the fact that quantum computers can do things "much faster" (in a sense that will be made precise later in this course) than classical computers. From this simple example it is also already clear that, while quantum systems can do multiple classical things at once, the tricky bit will be to somehow make use of that and extract all those multiple "computations" at the end.

1.2.2 Entanglement: connection without interaction

Entanglement is often only discussed once spin systems are introduced into quantum mechanics, but we have seen that entanglement really is a property of the quantum world that already exists for simple two-particle systems for which each particle is only labelled by a position. If the positions of the two particles are labelled by x_1 and x_2 respectively, then the wave function is some complex-valued function $\psi(x_1, x_2)$ of these two variables. In the special case that the function is separable, that is, when it takes the form

$$\psi(x_1, x_2) = \psi_1(x_1)\psi_2(x_2)\,, \tag{1.1}$$

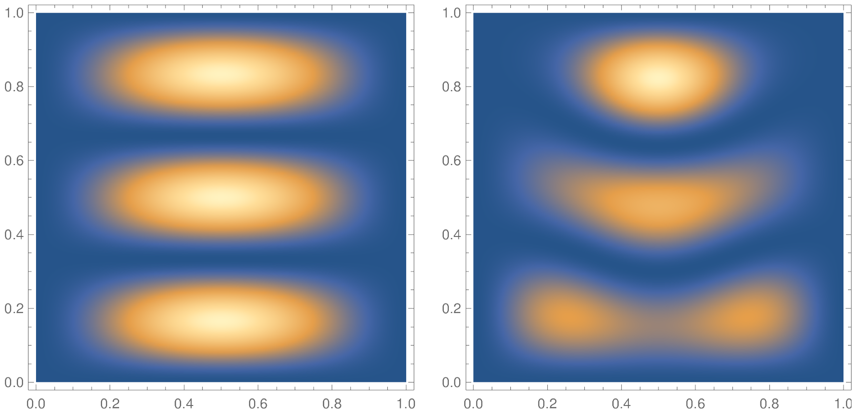


Figure 1.2: Probability density for two states of a two-particle system. On the left, the wave function is separable, while on the right it is non-separable.

the system is *non-entangled*. This means that a measurement of the position of particle 2 does not influence the measurement of the position of particle 1. If we do not measure particle 2, the probability density for particle 1 is

$$P_1(x_1) = \int \big|\psi(x_1, x_2)\big|^2 \mathrm{d}x_2 = \big|\psi_1(x_1)\big|^2\,, \tag{1.2}$$

because $\psi_2(x_2)$ is itself normalised. If we first measure the position of particle 2 to be $x_2 = q$, then the wave function collapses to

$$\tilde{\psi}(x_1) = N\psi(x_1, q) = N\psi_1(x_1)\psi_2(q) = \psi_1(x_1)\,, \tag{1.3}$$

where the constant N is determined by imposing that $\tilde{\psi}(x_1)$ is normalised. Because the wave function was separable, the probability density is the same as before, $P_1(x_1) = \big|\psi_1(x_1)\big|^2$. In terms of the picture above, the measurement of particle 1 before we know anything about particle 2 is obtained by integrating over the vertical direction, while measuring particle 1 after we measure particle 2 is done by evaluating the function at a particular vertical position. For a separable function, these are equivalent.

However, when the wave function is not separable, these two computations generally differ. For instance, assume that the wave function is

$$\psi(x_1, x_2) = \psi_1(x_1)\psi_2(x_2) + \chi_1(x_1)\chi_2(x_2) \tag{1.4}$$

for four different functions $\psi_1, \psi_2, \chi_1, \chi_2$. Integrating the complex norm-squared over x_2 or inserting a particular value $x_2 = q$ will now typically not produce the same function of x_1. In the picture on the right above, the difference is clear: *integrating* along the vertical direction x_2 does not produce the same function of x_1 as *slicing* the plot along some horizontal line $x_2 = q$.

Entangled states lead to one particle influencing the other even when there is no interaction, and even when the particles are separated by an arbitrary distance.

So typically, the measurement of the position of particle 2 *changes* the subsequent probability distribution for the measurement of the position of particle 1. This is what is known as *entanglement*. The important aspect of this is that measuring one particle immediately has consequences for the measurement of the other particle, even though the particles do not interact and even though the particles can be arbitrarily far separated. This never happens in classical mechanics.

Applications of entanglement: secure communication and quantum key distribution, teleportation, deciding whether nature is fundamentally non-classical (“Bell inequalities”).

Entanglement provides us with new ways to secure communication channels and also allows us to "teleport" a quantum state to a different location using classical communication. Moreover, it has played a crucial role in deciding whether the probabilistic character of quantum mechanics is merely due to "lack of understanding", or whether it is something fundamental. We will see that there are certain inequalities (the "Bell inequalities") which will tell us that nature is *not* classical.

1.3 Computers and information

Both examples discussed above are in a way much more complicated than the ones that we will be discussing at length in the present notes. For the purpose of quantum computing, we will no longer look at systems with a *continuous* space of classical states, but rather at much simpler ones in which the possible classical states are discrete. Think simply about ordinary digital computers: these are built from "bits", which are simple classical things which can take either one of two values.

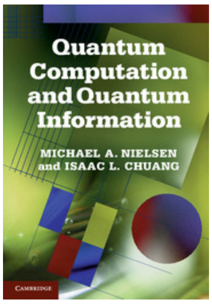
Application of quantum information: Von Neumann and Entanglement entropy.

Restricting to discrete systems will make our life a lot easier: there are no integrals but only sums, and all Hilbert spaces are finite-dimensional and operators acting in them can thus (if we want) be written out in terms of explicit finite-size matrices. With those systems, we can (and will, in the last chapter of these notes) develop the formalism to quantify how much information ("quantum entropy") is stored in a quantum system.

Where practical quantum computation mainly struggles at the moment is in *scaling* such systems up to respectable sizes. While typical digital computers, like your phone, contain at least on the order of 10^{10} to 10^{11} classical bits, the most powerful quantum computers around the time of writing do not get beyond 10^3 quantum bits. That these can still do useful things, and that scaling this up by only a few orders of magnitude will lead to dramatical changes in e.g. cryptography, is something that this module will try to get explain.

1.4 Recommended literature

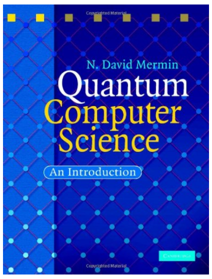
There is a large body of literature, both in book form an in the form of online courses, that deals with quantum information and quantum computing. The list below includes some of my own personal favourites, but that does of course not mean that other resources cannot be useful.


Quantum Computation and Quantum Information
MICHAEL A. NIELSEN and ISAAC L. CHUANG
CAMBRIDGE

- *Quantum Computation and Quantum Information*, Michael Nielsen and Isaac Chuang. This is a very big book (600 pages without counting the appendices), which I think is too much to get the key ideas across, but it does count as a bible in the field. If you have a question, it is probably answered in here somewhere.

- *Quantum computing for the very curious*, Andy Matuschak and Michael Nielsen. Shares one author with the book above. A modern online text with some new experimental techniques to make it easier to remember things. Content-wise, it's much shorter.

- *Quantum Computer Science*, N. David Mermin. A more recent book, essentially



only about quantum computing, and so mostly relevant for the Epiphany part of our module.

- *Quantum Computation*, John Preskill. Online notes for a course at Caltech, which for many people are one of the standard references in the field. Lucid explanation of the Bell inequalities (in chapter 4), which we will discuss later this term.

- *Basic Quantum Algorithms*, Renato Portugal. Very clear exposition of the most basic quantum algorithms, preceded by a quick introduction to elements of quantum information and quantum computing.

If you do not care about any of the physics background, and just want to “get going” with the maths of quantum information, there are some more recent resources which can help:

- *Basics of quantum information*, a course provided by IBM and the authors of the Qiskit software. There are video lectures by John Watrous and there is an online/pdf written set of lectures notes as well.

2

Quantum mechanics essentials

LECTURE 2

Quantum mechanics is commonly introduced by discussing the concept of a *wave function*. This is a function $\psi(x)$ such that the complex norm $|\psi(x)|^2$ gives the probability of finding the system in state x. We have seen that the space of wave functions can be seen as an infinite-dimensional vector space, called *Hilbert space*. Here we will revise these concepts, and then introduce the *Dirac notation*, which is more compact and also more suitable to describe systems which classically have only a discrete number of states.

2.1 States and wave functions

Let us reminder ourselves of the quantum mechanics of a single particle on the real line $x \in \mathbb{R}$. In the formulation of quantum mechanics using wave functions, this system is described by a complex-valued wave function of space and time, $\psi(x,t)$. The probability of finding a particle in a region $x \in [a,b]$ at some moment in time t is given by

$$P(a,b;t) = \int_a^b \left|\psi(x,t)\right|^2 \mathrm{d}x\,. \tag{2.1}$$

This probability is clearly ≥ 0 because of the properties of the norm of a complex number. We normalise the wave function such that

$$P(-\infty,+\infty;t) = 1\,, \tag{2.2}$$

that is to say, the probability of finding the particle somewhere is one.

The time-evolution of the wave function is given by the Schrödinger equation,

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t) + V(x)\psi(x,t)\,. \tag{2.3}$$

The right-hand side equals the Hamiltonian operator acting on the wave function, $\hat{H}\psi(x,t) = \hat{K}\psi(x,t) + \hat{V}\psi(x,t)$ where K, V are the kinetic and potential energy respectively. If we impose (2.2) for a single moment in time $t = t_0$, then the Schrödinger equation guarantees that it will remain valid for arbitrary other times.

The function $\psi(x)$ is said to describe the *state* of the system. Rather than knowing exactly where a particle is, as we do classically, we only know the probability density $P(x)$ (or more precisely, the amplitude $\psi(x)$) of finding the particle somewhere on the real line. Because the Schrödinger equation is *linear*, any linear superposition of wave functions is also a solution. So you can have a wave function strongly peaked on earth, and another one strongly peaked at the moon, and the linear combination is still a valid quantum mechanical wave function,

$$\psi(x) = \psi_{\text{earth}}(x) + \psi_{\text{moon}}(x)\,. \tag{2.4}$$

In quantum mechanics, only measurement will force the system into one of the two classical configurations.

2.2 The Dirac notation (bra-ket)

Because the space of wave functions is linear (wave functions can be superposed) and because we have a norm, we can view each wave function as a *vector* in a (complex) vector space: *Hilbert space*. Typically, this vector space will be infinite-dimensional, because there is an infinite, or even a continuum, of possible classical configurations. But this is not necessary, and in fact we will consider in this module mainly systems for which we have only a finite (but possibly large) number of classical configurations.

In Dirac notation, the ket $|\psi\rangle \in \mathcal{H}$ corresponds to the physical state. The dual bra vectors $\langle\phi|$ live in $\mathcal{H}^*$ and together they can form the inner product $\langle\phi|\psi\rangle$.

The Dirac notation consists in writing $|\psi\rangle$ for a vector in Hilbert space corresponding to the wave function $\psi(x)$. And instead of calling it a vector, we call it a *ket*, for reasons that will become clear shortly. The fact that we no longer write the x label is significant. Compare the situation in linear algebra. There, we can have a physical, arrow-like object which we call a vectors (let's say $\mathbf{v}$). To write down concretely which vector we mean, we choose a basis of unit vectors, and then write down the *components* of the vector on that basis, e.g. $\mathbf{v} = (2, 3)$. But changing the basis does not change the vector itself, only its components. With wave functions a similar thing happens. The representation $\psi(x)$ refers to the "basis" of position eigenstates labelled by the position x. But it is perfectly possible to write down the wave function in a different basis, for instance the basis of momentum eigenstates.

So we use $|\psi\rangle$ from now on, as a more abstract way of expressing the vector in Hilbert space. For any two such vectors, we have a positive definite inner product for the corresponding wave fuctions,

$$\text{inner product}(\phi, \psi) = \int_{-\infty}^{\infty} \phi^*(x,t)\psi(x,t)\,\mathrm{d}x =: \langle\phi|\psi\rangle\,. \tag{2.5}$$

On the right-hand side we have introduced the inner product in the Dirac notation, $\langle\phi|\psi\rangle$. It requires that we have access to the *dual* vector $\langle\phi|$, which as you can see from the explicit integral representation, is simply related to the complex conjugate of the wave function ϕ. This new object $\langle\phi|$ is called a *bra*, so that the inner product (or bracket) between two states reads bra-ket.

In this module we will almost always consider finite-dimensional Hilbert spaces. That means that we can define the Hilbert space by specifying a finite set of basis states. Of course, there is not a unique choice of basis states, and it will often be useful to consider different choices (related by changes of basis matrices). Commonly we will consider orthonormal bases, i.e. ones where all the basis states have norm 1 and are mutually orthogonal.

To better understand the bra symbols we need to introduce the concept of *dual* of a vector space V. Formally, the dual V^* of a vector space V is the vector space of linear functionals from V into $\mathbb{C}$. Or in formulas,

$$V^* = \{\Phi : V \to \mathbb{C} \text{ s.t. } \Phi(a\mathbf{z} + b\mathbf{w}) = a\,\Phi(\mathbf{z}) + b\,\Phi(\mathbf{w}),\\ \forall\, a\,, b \in \mathbb{C}\,,\ \text{and}\,\forall\,\mathbf{z}\,,\mathbf{w} \in V\}\,. \tag{2.6}$$

> **Exercise:**
>
> Show that this is indeed a vector space over the complex numbers and show that if V has dimension n so does V^*.

But this dual space is nothing mysterious. If $V = \mathbb{C}^n$, vectors in the standard basis are simple n dimensional column vectors, i.e. $n \times 1$ matrices if you wish, similarly you can think of V^* as the vector space of $1 \times n$ matrices, i.e. row vectors. Furthermore whenever V is endowed with a complex inner product $\langle \cdot, \cdot \rangle$, precisely as our Hilbert space of states $\mathcal{H}$, for each vector $\mathbf{z} \in V$ we can associate an element $\Phi_{\mathbf{z}} \in V^*$ schematically as $\Phi_{\mathbf{z}} = \langle \mathbf{z}, \cdot \rangle$

$$\Phi_{\mathbf{z}}(\mathbf{w}) \doteq \langle \mathbf{z}, \mathbf{w} \rangle$$

for all $\mathbf{w} \in V$. It is easy to check (just write down what it means) that $\Phi_{\mathbf{z}}(\cdot)$ just defined is indeed a linear functional from V into $\mathbb{C}$, hence $\Phi_{\mathbf{z}} \in V^*$. This means that as soon as our vector space V has a complex inner product we have immediately an isomorphism between V and V^*

$$V \ni \mathbf{z} \mapsto \Phi_{\mathbf{z}}(\cdot) = \langle \mathbf{z}, \cdot \rangle \, \in V^* \, .$$

In quantum mechanics we have vectors, i.e. elements of $\mathcal{H}$, denoted by ket vectors $|\psi\rangle$, and thanks to the complex inner product for each vector $|\psi\rangle$ we can consider the corresponding element in the dual space $\langle\psi| \in \mathcal{H}^*$. A bra vector $\langle\psi|$ applied to a ket vector $|\phi\rangle$ gives precisely the inner product $\langle \psi \,|\phi\rangle$.

A ket $|\psi\rangle$ can be thought of as a column-vector, and then the bra $\langle\phi|$ is a row vector, so that their inner product is a scalar.

Note in particular that it does **NOT** make any sense to consider $|\psi\rangle + \langle\phi|$ since one is a column vector that cannot be added to the other object which is a row vector!

> **Example:**
>
> Suppose our quantum mechanical system is described by a three-dimensional Hilbert space $\mathcal{H}$ written in terms of the orthonormal basis $|0\rangle, |1\rangle, |2\rangle$, i.e. $\mathcal{H} = \mathrm{span}\{|0\rangle, |1\rangle, |2\rangle\}$. We can represent any vector in $\mathcal{H}$ as a three dimensional column vector using the standard basis
>
> $$\begin{aligned} &|0\rangle \mapsto \begin{pmatrix}1\\0\\0\end{pmatrix}, |1\rangle \mapsto \begin{pmatrix}0\\1\\0\end{pmatrix}, |2\rangle \mapsto \begin{pmatrix}0\\0\\1\end{pmatrix}, \\ &|\psi\rangle = a\,|0\rangle + b\,|1\rangle + c\,|2\rangle \mapsto \begin{pmatrix}a\\b\\c\end{pmatrix}, \\ &\langle\psi| = a^*\,\langle 0| + b^*\,\langle 1| + c^*\,\langle 2| \\ &\quad \mapsto \begin{pmatrix}a^* & b^* & c^*\end{pmatrix} = \begin{pmatrix}a\\b\\c\end{pmatrix}^\dagger . \end{aligned}$$
>
> To understand the second line let us remember that we are told that three basis vectors are orthonormal hence we know that the matrix that represents the inner product in this basis is given by the identity matrix. At this point it is very simple to compute the inner product between two states, say the inner product of $|\phi\rangle =$

> $d\left|0\right\rangle + e\left|1\right\rangle + f\left|2\right\rangle$ with $\left|\psi\right\rangle = a\left|0\right\rangle + b\left|1\right\rangle + c\left|2\right\rangle$
>
> $$\left\langle\phi\right|\psi\rangle = \begin{pmatrix} d^* & e^* & f^* \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} d \\ e \\ f \end{pmatrix}^{\dagger}\begin{pmatrix} a \\ b \\ c \end{pmatrix},$$
>
> where remember $A^\dagger = (A^*)^T$ is the transpose complex conjugate.
>
> In particular we also see that if the ket $\left|\psi\right\rangle = a\left|0\right\rangle + b\left|1\right\rangle + c\left|2\right\rangle$ is represented by the column vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$, then the bra $\left\langle\psi\right|$ can really be thought of as $\left\langle\psi\right| = (\left|\psi\right\rangle)^\dagger$ and represented by the row vector $(a, b, c)^\dagger = (a^*, b^*, c^*)$.

2.3 Hilbert space formalities

You can now write this all up in formal language if you want. A quantum mechanical system is described by a ket $\left|\psi\right\rangle$ in Hilbert space $\mathcal{H}$. A Hilbert space is a (complex) vector space with Hermitian inner product. This means that for any $\left|\psi\right\rangle \in \mathcal{H}$ and $\left|\phi\right\rangle \in \mathcal{H}$:

- For any complex numbers a and b, $(a\left|\psi\right\rangle + b\left|\phi\right\rangle) \in \mathcal{H}$.

  (*linear combinations of vectors = quantum superposition*)

- The inner product of $\left|\psi\right\rangle$ with $\left|\phi\right\rangle$ is a complex number denoted

  $$\left\langle\psi\right|\phi\rangle = \left\langle\phi\right|\psi\rangle^* \in \mathbb{C}.$$

  The inner product is Hermitean,

  $$\left\langle\psi\right|\phi\rangle = \left(\left\langle\phi\right|\psi\rangle\right)^*.$$

- The inner product is linear in the second state (and so anti-linear in the first state). I.e. if $\left|\phi\right\rangle = c_1\left|\phi_1\right\rangle + c_2\left|\phi_2\right\rangle$ then

  $$\begin{aligned} \left\langle\psi\right|\phi\rangle &= c_1\left\langle\psi\right|\phi_1\rangle + c_2\left\langle\psi\right|\phi_2\rangle\ , \\ \left\langle\phi\right|\psi\rangle &= c_1^*\left\langle\phi_1\right|\psi\rangle + c_2^*\left\langle\phi_2\right|\psi\rangle\ . \end{aligned}$$

  In other words, the inner product is linear in the *second* factor, and anti-linear in the *first*; it is *sesquilinear*.

  Note that in your linear algebra module you might have seen a slightly different definition for an hermitian inner product which is linear in the **first** term! This is just a convention and in this module we will keep the inner product to be linear in the second term. Combining linearity in the second term with hermiticity tells us that the inner product is not quite linear in the first term

- This inner product is real, $\left\langle\psi\right|\psi\rangle \in \mathbb{R}$. However, we also have a *physical state* condition (and we will only consider such states in this module): $\left\langle\psi\right|\psi\rangle \geq 0$ and $\left\langle\psi\right|\psi\rangle = 0 \iff \left|\psi\right\rangle = 0$. We will use the notation $||\left|\psi\right\rangle|| \equiv \sqrt{\left\langle\psi\right|\psi\rangle}$ for the *norm* of $\left|\psi\right\rangle$.

Finally, states which differ only by a normalisation factor are physically equivalent, i.e.

$$|\psi\rangle \sim c\,|\psi\rangle$$

for any non-zero $c \in \mathbb{C}$. There are two ways to work with this equivalence relation. One is to ignore the normalisation but then include appropriate factors of the norms of states in formulae. The other, which we will usually assume, is to always work with *normalised states*, i.e. unless indicated otherwise a state $|\psi\rangle$ will be assumed to be have $||\,|\psi\rangle\,|| = 1$. If you have a state which is not normalised, just divide it by its norm to get a normalised state. Note that normalisation does not fix a unique representative of the equivalence class of states since multiplying by a phase $\exp(i\theta)$ for any real phase θ does not change the norm, i.e. $||\,|\psi\rangle\,|| = 1$ if and only if $||e^{i\theta}\,|\psi\rangle\,|| = 1$.

Physical states are *rays* in Hilbert space, as the normalisation is irrelevant (and often tuned so the norm of the vector is one).

Sometimes (pure) quantum mechanical states are called *rays* in the Hilbert space because of the equivalence $|\psi\rangle \sim c\,|\psi\rangle$ with $c \in \mathbb{C}$ non-zero.

NOTE: Obviously the zero state cannot be normalised but that is OK as it does not describe the state of a physical system, and there is no physical process to transform a non-zero state to the zero state[1].

2.4 Operators

In quantum mechanics we work with linear operators acting on the states in a Hilbert space. Such operators are used to describe the time-evolution of the system and to describe measurements. If $\hat{A}$ is a linear operator then acting on linear combinations of states we have

$$\hat{A}(a\,|\psi\rangle + b\,|\phi\rangle) = a(\hat{A}\,|\psi\rangle) + b(\hat{A}\,|\phi\rangle)\,,$$

i.e. it is linear. Also, products and linear combinations of linear operators, are again linear operators.

The *adjoint* (also commonly called the Hermitian conjugate) of $\hat{A}$ is denoted $\hat{A}^\dagger$ and defined by

$$\left\langle\psi\right|\left(\hat{A}^\dagger|\phi\right\rangle\right) = \left[\left\langle\phi\right|\left(\hat{A}|\psi\right\rangle\right)\right]^*$$

for all states $|\psi\rangle$ and $|\phi\rangle$.

In quantum mechanics, we usually focus on two types of linear operators:

- *self-adjoint operators* (or Hermitian) meaning $\hat{H}^\dagger = \hat{H}$. Self-adjoint operators correspond to observables, i.e. quantities which can be measured. E.g. $\hat{X}$ position operator, $\hat{P}$ momentum operator, $\hat{H}$ Hamiltonian operator, $\hat{S}$ spin operator. The reason for that comes from the fact that hermitian operators have real eigenvalues.

- *unitary operators* meaning $\hat{U}^\dagger\hat{U} = \hat{I}$. Unitary operators are used to describe time-evolution in quantum mechanics.

[1] Do not confuse the zero state (meaning the unique state with norm zero) with a state labelled by zero, i.e. $|0\rangle \neq 0$! The norm of $|0\rangle$ is always non-zero $||\,|0\rangle\,|| = 1$, while the norm of the 0 vector is always vanishing $||0|| = 0$.

> **Exercise:**
>
> Show that the eigenvalues of a self-adjoint operator $\hat{H}$ must be real. Show that eigenstates of a self-adjoint operator corresponding to different eigenvalues are automatically orthogonal to each others.

Note: when we pass to matrices: self-adjoint operators become hermitean matrices, i.e. $H^\dagger = (H^T)^\star = H$ while unitary operators become unitary matrices, i.e. $U^\dagger U = UU^\dagger = I$. Do not confuse these two properties!

> **Exercise:**
>
> Given the matrices
>
> $$\begin{aligned} A &= \begin{pmatrix} 1 & i \\ -i & 2 \end{pmatrix}\,, \\ B &= \begin{pmatrix} \cos(\sqrt{5}) + \frac{2i\sin(\sqrt{5})}{\sqrt{5}} & \frac{\sin(\sqrt{5})}{\sqrt{5}} \\ -\frac{\sin(\sqrt{5})}{\sqrt{5}} & \cos(\sqrt{5}) - \frac{2i\sin(\sqrt{5})}{\sqrt{5}} \end{pmatrix}\,, \\ C &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,, \\ D &= \begin{pmatrix} 1 & i \\ 1+i & 3 \end{pmatrix}\,, \end{aligned}$$
>
> check that A is hermitean, B is unitary, C is both unitary and hermitean, D is neither unitary nor hermitean.

> **Exercise:**
>
> Find a basis for the vector space (check that it is indeed a vector space!) of the 2×2 Hermitean matrices.

We also define the commutator of two operators $\hat{A}$ with $\hat{B}$ as:

$$\left[\hat{A}, \hat{B}\right] \equiv \hat{A}\hat{B} - \hat{B}\hat{A}.$$

There is also a similar definition of the anti-commutator

$$\left\{\hat{A}, \hat{B}\right\} \equiv \hat{A}\hat{B} + \hat{B}\hat{A}.$$

The *expectation value* of an observable $\hat{A}$ on a state ψ denoted by $\langle A \rangle_\psi$ is given by

$$\langle A \rangle_\psi = \langle \psi | \, \hat{A} \, | \psi \rangle \ .$$

In most cases if there is no confusion regarding which state we are considering we will drop the subscript and simply write $\langle A \rangle$. This expectation value can really be interpreted as the average outcome of many measurements of the same observable $\hat{A}$ on the same state $|\psi\rangle$, i.e. prepare 1000 times the same state $|\psi\rangle$, measure 1000 times the same observable $\hat{A}$ and then take the average.

The expectation value of an operator gives the average outcome of the measurement of the corresponding observable, if we start with the same system many times over.

Note that the expectation value is clearly a real number

$$\begin{aligned} \langle A \rangle_\psi^* &= (\langle \psi | \, \hat{A} \, | \psi \rangle)^* = (\langle \psi | \, \hat{A}^\dagger \, | \psi \rangle) \\ &= (\langle \psi | \, \hat{A} \, | \psi \rangle) = \langle A \rangle_\psi \,, \end{aligned} \tag{2.7}$$

where the hermiticity of the inner product, i.e. $\langle \psi \, | \phi \rangle^* = \langle \phi \, | \psi \rangle$, and the hermiticity of $\hat{A}$, $\hat{A}^\dagger = \hat{A}$, both play a crucial role.

2.5 Matrix representation

LECTURE 3

As already mentioned above, for an N-dimensional Hilbert space, we can represent states using ket vectors or complex N-component column vectors, similarly for bra vectors we need N-component row vectors. A standard choice is to represent the basis states by column vectors with all components zero, except a single 'one'. Then, provided we are using an orthonormal basis (so that the matrix that represents the inner product is given just by the identity matrix), the inner product of two states represented by column vectors $\mathbf{u}$ and $\mathbf{v}$ is given by standard matrix multiplication as $\mathbf{u}^\dagger\mathbf{v}$. Linear operators are then represented by $N \times N$ matrices.

By writing out the action of an operator on each of the basis vectors in Hilbert space, we can construct its *matrix representation*.

In such a representation self-adjoint operators $\hat{H}^\dagger = \hat{H}$ are indeed Hermitian matrices $H^\dagger = H$, and unitary operators $\hat{U}^\dagger\hat{U} = \hat{I}$ are unitary matrices $U^\dagger U = I$ where I is the identity matrix of the appropriate dimension. Note that in most cases we will keep the hat symbol $\hat{}$, as in $\hat{A}$, to denote the abstract operator without having having picked any particular basis to be represented as the standard one, once we choose a particular orthonormal basis to be represented via the standard one we will refer to the matrix representing the operator $\hat{A}$ in this basis with the same letter but without the hat, i.e. just A. Note that we will be playing a lot with different basis so although the abstract operator is one $\hat{A}$ it might be represented by different matrices A_1 , A_2 , ... according to which basis we choose! However we should remember from Linear Algebra I that if we change basis, say from $\{\mathbf{v}_1, ...\mathbf{v}_2\}$ to $\{\mathbf{w}_1, ..., \mathbf{w}_n\}$, then the matrix A_2 representing the linear transformation $\hat{A}$ in the new basis is related to the matrix A_1 representing the same linear transformation $\hat{A}$ but in the old basis via

$$A_2 = S^{-1} A_1 S\,,$$

where S is the change of basis matrix to go from the new basis to the old one.

> **Example:**
>
> Let $\mathcal{H} = \text{span}\{|0\rangle\,, |1\rangle\}$ be a 2-dimensional Hilbert space of states, and assume that the basis vector are orthonormal. We are given a linear operator $\hat{A}$ defined on this basis
>
> $$\hat{A}\,|0\rangle = a\,|0\rangle + b\,|1\rangle\,, \qquad \hat{A}\,|1\rangle = c\,|0\rangle + d\,|1\rangle\,, \tag{2.8}$$
>
> with $a, b, c, d \in \mathbb{C}$.
>
> First of all we can pass to the vector/matrix representation by using coordinates. We chose to represent the first basis vector $|0\rangle \to \begin{pmatrix}1\\0\end{pmatrix}$ and the second one as $|1\rangle \to \begin{pmatrix}0\\1\end{pmatrix}$. Once we make this choice of basis the operator $\hat{A}$ can be represented as the 2×2 matrix
>
> $$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}\,,$$
>
> and the abstract form $\hat{A}\,|0\rangle = a\,|0\rangle + b\,|1\rangle$ can be simply stated in matrix language as $A\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}a\\b\end{pmatrix}$, and similarly for the other basis vector.
>
> Let us now compute the adjoint $\hat{A}^\dagger$ of $\hat{A}$ and find which conditions we have to impose on the coefficients a, b, c, d such that $\hat{A}$ becomes self-adjoint. If we want to

compute $\hat{A}^\dagger$ we need to know what this operator does on a basis, i.e. we need to compute $\hat{A}^\dagger \left|0\right\rangle$ and $\hat{A}^\dagger \left|1\right\rangle$.

So we can write

$$\hat{A}^\dagger \left|0\right\rangle = \alpha \left|0\right\rangle + \beta \left|1\right\rangle \ , \qquad \hat{A}^\dagger \left|1\right\rangle = \gamma \left|0\right\rangle + \delta \left|1\right\rangle \ ,$$

for some yet undetermined $\alpha, \beta, \gamma, \delta \in \mathbb{C}$. To fix these coefficients we need to remember that $\left|0\right\rangle, \left|1\right\rangle$ form an orthonormal basis so we have

$$\begin{aligned}
\left\langle 0\right| \hat{A}^\dagger \left|0\right\rangle &= \left\langle 0\right| (\alpha \left|0\right\rangle + \beta \left|1\right\rangle) = \alpha\,, \\
\left\langle 1\right| \hat{A}^\dagger \left|0\right\rangle &= \left\langle 1\right| (\alpha \left|0\right\rangle + \beta \left|1\right\rangle) = \beta\,, \\
\left\langle 0\right| \hat{A}^\dagger \left|1\right\rangle &= \left\langle 0\right| (\gamma \left|0\right\rangle + \delta \left|1\right\rangle) = \gamma\,, \\
\left\langle 1\right| \hat{A}^\dagger \left|1\right\rangle &= \left\langle 1\right| (\gamma \left|0\right\rangle + \delta \left|1\right\rangle) = \delta\,.
\end{aligned}$$

Finally we need to remember the definition $\left\langle \psi \middle| \hat{A}^\dagger \middle| \phi \right\rangle = \left\langle \phi \middle| \hat{A} \middle| \psi \right\rangle^*$ so we have

$$\begin{aligned}
\alpha &= \left\langle 0\right| \hat{A}^\dagger \left|0\right\rangle = (\left\langle 0\right| \hat{A} \left|0\right\rangle)^* = a^*\,, \\
\beta &= \left\langle 1\right| \hat{A}^\dagger \left|0\right\rangle (\left\langle 0\right| \hat{A} \left|1\right\rangle)^* = c^*\,, \\
\gamma &= \left\langle 0\right| \hat{A}^\dagger \left|1\right\rangle = (\left\langle 1\right| \hat{A} \left|0\right\rangle)^* = b^*\,, \\
\delta &= \left\langle 1\right| \hat{A}^\dagger \left|1\right\rangle = (\left\langle 1\right| \hat{A} \left|1\right\rangle)^* = d^*\,.
\end{aligned}$$

Note the order of the vectors being flipped! We have then

$$\hat{A}^\dagger \left|0\right\rangle = a^* \left|0\right\rangle + c^* \left|1\right\rangle \ , \qquad \hat{A}^\dagger \left|1\right\rangle = b^* \left|0\right\rangle + d^* \left|1\right\rangle \ .$$

Using the same basis as above this operator can be represented as the 2×2 matrix

$$A^\dagger = \begin{pmatrix} a^* & b^* \\ c^* & d^* \end{pmatrix} \ .$$

When passing to coordinates the matrix representing the adjoint operator $\hat{A}^\dagger$ is exactly $A^\dagger = (A^*)^T$, i.e. the transpose complex conjugate of the matrix A representing the operator $\hat{A}$. Finally if we want the operator to be self-adjoint we must have $\hat{A}^\dagger = \hat{A}$ which imposes $a = a^*$, $b = c^*$, $d = d^*$. These conditions are identical to imposing that the matrix representing $\hat{A}$ is an hermitian matrix, i.e. $A = A^\dagger$.

Once we realise we are just doing linear algebra we can easily understand what happens when we change basis. Suppose for example we are given the operator

$$\hat{B} \left|0\right\rangle = 2i \left|0\right\rangle + 5 \left|1\right\rangle \ , \qquad \hat{B} \left|1\right\rangle = -3 \left|0\right\rangle + (1+i) \left|1\right\rangle \ ,$$

which can be represented in the same basis as above by the matrix $B = \begin{pmatrix} 2i & -3 \\ 5 & 1+i \end{pmatrix}$.

We want to use now a new orthonormal basis defined by $\left|\pm\right\rangle = \frac{1}{\sqrt{2}}(\left|0\right\rangle \pm \left|1\right\rangle)$ (check that this is indeed an orthonormal basis) or equivalently $\left|0\right\rangle = \frac{1}{\sqrt{2}}(\left|+\right\rangle + \left|-\right\rangle)$, $\left|1\right\rangle = \frac{1}{\sqrt{2}}(\left|+\right\rangle - \left|-\right\rangle)$. We can proceed in two ways. One possibility is to rewrite the action of $\hat{B}$ in this new basis

$$\begin{aligned}
\hat{B}\left|+\right\rangle &= \frac{1}{\sqrt{2}}\hat{B}(\left|0\right\rangle + \left|1\right\rangle) \\
&= \frac{1}{\sqrt{2}}((-3+2i)\left|0\right\rangle + (6+i)\left|1\right\rangle) \\
&= \frac{(-3+2i)}{2}(\left|+\right\rangle + \left|-\right\rangle) + \frac{(6+i)}{2}(\left|+\right\rangle - \left|-\right\rangle) \\
&= \frac{3+3i}{2}\left|+\right\rangle + \frac{-9+i}{2}\left|-\right\rangle \,, \\
\hat{B}\left|-\right\rangle &= \frac{1}{\sqrt{2}}\hat{B}(\left|0\right\rangle - \left|1\right\rangle) \\
&= \frac{1}{\sqrt{2}}((3+2i)\left|0\right\rangle + (4-i)\left|1\right\rangle) \\
&= \frac{(3+2i)}{2}(\left|+\right\rangle + \left|-\right\rangle) + \frac{(4-i)}{2}(\left|+\right\rangle - \left|-\right\rangle) \\
&= \frac{7+i}{2}\left|+\right\rangle + \frac{-1+3i}{2}\left|-\right\rangle \,.
\end{aligned}$$

Hence in the new basis represented by $\left|+\right\rangle \to \begin{pmatrix}1\\0\end{pmatrix}$, $\left|-\right\rangle \to \begin{pmatrix}0\\1\end{pmatrix}$ the new matrix $\tilde{B}$ representing the *same* operator $\hat{B}$ now takes the form

$$\tilde{B} = \frac{1}{2}\begin{pmatrix}3+3i & 7+i\\ -9+i & -1+3i\end{pmatrix} \,.$$

We could have reached the same conclusion noting that we are just making a change of basis from $\{\left|0\right\rangle, \left|1\right\rangle\}$ to $\{\left|+\right\rangle, \left|-\right\rangle\}$ and the change of basis matrix is simply given by

$$S = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \,,$$

hence the matrix representation $\tilde{B}$ of the same operator $\hat{B}$ but in the new basis, is simply given by $S^{-1}BS$ where B is the matrix representation of $\hat{B}$ in the old basis. This matrix multiplication produces exactly the same matrix $\tilde{B}$ just computed above.

Change of basis will play a crucial role in the discussion of qubits. We will have a set of privileged operators and we will keep on changing from a basis of eigenvectors for one such operator to a basis of eigenvectors for another of these operators. Every time we change basis the matrix representing these operators will change according to a change of basis transformation, i.e. $S^{-1}BS$.

2.6 Time-evolution

In quantum mechanics the time-evolution of the system is governed by a self-adjoint operator called the Hamiltonian $\hat{H}$. In this module we work in the Schrödinger picture, so a state in the system evolves with time, and we can consider a state $|\psi(t)\rangle$ (think of it as a time-dependent vector). The time-evolution is described by the Schrödinger equation

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle \, .$$

This can also be written in an integrated form to define the state in terms of some initial state, say at time $t > 0$ in terms of the state at $t = 0$:

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

where $\hat{U}_t$ is a unitary operator. In the case where the Hamiltonian operator is not time-dependent we have

$$\hat{U}_t = \exp\left(-\frac{i}{\hbar}t\,\hat{H}\right) .$$

In quantum information we usually assume complete control over a quantum system or subsystem. This means that we can interact with the system in a arbitrary way, e.g. by rotating it, applying electric or magnetic fields etc. In terms of time evolution this means that we assume we have the ability to transform the state of the system

$$|\psi\rangle \to \hat{U}|\psi\rangle$$

using any unitary operator $\hat{U}$ we want. As such we will usually talk about transformations by a unitary operator $\hat{U}$, rather than in terms of a Hamiltonian operator with evolution for some specific period of time.

2.6.1 Exponential of operators

In this Section we defined the time evolution operator in terms of the exponential of the Hamiltonian operator. This is a general concept:

Def: The exponential of a matrix A, or more generally of an operator $\hat{A}$, is defined by the Taylor series

$$\exp(\hat{A}) = \sum_{n=0}^{\infty}\frac{\hat{A}^n}{n!} = \hat{I} + \frac{\hat{A}}{1!} + \frac{\hat{A}^2}{2!} + \dots ,$$

where $\hat{I}$ denotes the identity operator. Note that for the operators we will consider this series will always converge.

> **Example:**
>
> Suppose $A = \mathrm{diag}(\lambda_1, ...\lambda_N)$ be a diagonal $N \times N$ matrix with $\lambda_i \in \mathbb{C}$. Let us compute e^A. To this end we need to compute A^n which is $A \cdot A \cdot ... \cdot A$ n-times. In general this is a difficult task but for A diagonal it is actually very simple $A^n =$

diag$(\lambda_1^n,...\lambda_N^n)$ hence

$$\begin{aligned}e^A &= \sum_{n=0}^\infty \frac{A^n}{n!} = \sum_{n=0}^\infty \frac{1}{n!}\text{diag}(\lambda_1^n,...\lambda_N^n)\\ &= \text{diag}(\sum_{n=0}^\infty \frac{\lambda_1^n}{n!}, \sum_{n=0}^\infty \frac{\lambda_2^n}{n!},..., \sum_{n=0}^\infty \frac{\lambda_N^n}{n!})\\ &= \text{diag}(e^{\lambda_1},...,e^{\lambda_N})\,.\end{aligned}$$

So e^A is once again diagonal with diagonal elements simply given by the exponential of the diagonal elements of A.

Example:

Consider

$$N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

and compute e^{tN} with $t \in \mathbb{R}$. First we notice that N is a nilpotent matrix, i.e. we can find $m \in \mathbb{N}$ such that $N^m = 0$, in particular in this case

$$N^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and $N^3 = 0$, i.e. the 3×3 zero matrix. In this case the exponential series truncates after finitely many terms

$$e^{tN} = \sum_{n=0}^\infty \frac{t^n N^n}{n!} = \mathbb{I}_3 + tN + \frac{t^2N^2}{2!} = \begin{pmatrix} 1 & t & \frac{t^2}{2} \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}\,.$$

Note in particular that e^{tN} is **NOT** simply given by the exponential of each entries of tN!

Example:

Let $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, we want to compute $U_2(\alpha) = e^{i\alpha\sigma_2}$ and show that this is a unitary matrix for every $\alpha \in \mathbb{R}$. First we notice that $(i\sigma_2)^2 = -\mathbb{I}_2$ hence $(i\sigma_2)^{2n} = (-1)^n\mathbb{I}_2$ while $(i\sigma_2)^{2n+1} = (-1)^n(i\sigma_2)$.

This alternating pattern between even and odd powers allows us to evalute the exponential

$$\begin{aligned}U_2(\alpha) = e^{i\alpha\sigma_2} &= \sum_{n=0}^\infty \frac{\alpha^n(i\sigma_2)^n}{n!}\\ &= \sum_{n\text{ even}} + \sum_{n\text{ odd}} = \sum_{n=0}^\infty \frac{(-1)^n\alpha^{2n}}{(2n)!}\mathbb{I}_2 + \sum_{n=0}^\infty \frac{(-1)^n\alpha^{2n+1}}{(2n+1)!}(i\sigma_2)\\ &= \cos(\alpha)\mathbb{I}_2 + i\sigma_2\sin(\alpha) = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}\,.\end{aligned}$$

> In the last line we have used the Taylor expansion for the sine and cosine functions
>
> $$\begin{aligned}\cos(\alpha) &= \sum_{n=0}^{\infty} \frac{(-1)^n \alpha^{2n}}{(2n)!}\,, \\ \sin(\alpha) &= \sum_{n=0}^{\infty} \frac{(-1)^n \alpha^{2n+1}}{(2n+1)!}\,.\end{aligned}$$
>
> Finally it is easy to check that $U_2(\alpha)^\dagger U_2(\alpha) = U_2(\alpha) U_2(\alpha)^\dagger = \mathbb{1}_2$ so the matrix $U_2(\alpha)$ is indeed a unitary matrix and hence a possible time evolution operator of a 2-dimensional system. We will see that this type of unitary time evolution will be crucial for the study of the qubit systems later on.

A final comment: From these examples it should be clear that in general to compute the exponential of a matrix you **CANNOT** simply compute the exponential of each entry! For more on exponentials of matrices and the various ways to compute them, see e.g. [4].

3

Measurement and uncertainty

3.1 Observables

LECTURE 4

In quantum mechanics observables are used to indicate quantities which could be measured in an experiment, and *observable* also refers to the self-adjoint operator associated to such a measurement. Specifically, there is a one-to-one correspondence between measurable quantities M and self-adjoint operators $\hat{M}$. One example is energy and the Hamiltonian operator $\hat{H}$.

Now, in quantum mechanics the possible values of a measurement of M are the eigenvalues of $\hat{M}$ (ignoring experimental error, we do not commit experimental errors.) Typically for a given state $|\psi\rangle$ we cannot predict with certainty the result of a measurement. instead we can give probabilities for the different outcomes. Note that for a Hilbert space $\mathcal{H}$ of finite dimension N, the operator $\hat{M}$ will have a finite number of eigenvalues, in fact the number is precisely N (if we count the multiplicity of any degenerate eigenvalues) since this is equivalent to asking for the spectrum of an $N \times N$ Hermitian matrix.

Observables correspond to Hermitian matrices. Their eigenvalues are the possible measurement outcomes.

Dictionary Linear Algebra $\leftrightarrow$ QM:

- Self-adjoint operators $\hat{H}$ $\leftrightarrow$ Quantum mechanical observables;
- Eigenvalues of a self-adjoint operator $\hat{H}$ $\leftrightarrow$ Possible outcomes of measuring that quantum mechanical observable; (Prove that the eigenvalues of a self-adjoint operator must be real numbers. You will never measure a complex outcome in a lab!)
- Eigenstates of a self-adjoint operator $|\psi_E\rangle$, i.e. $\hat{H}\,|\psi_E\rangle = E\,|\psi_E\rangle$ for some real eigenvalue $E \in \mathbb{R}$ $\leftrightarrow$ States of definite outcome. If you measure $\hat{H}$ on $|\psi_E\rangle$ with probability $p = 1$ you will find outcome E, the corresponding eigenvalue.

Def: The *spectrum* of an operator $\hat{H}$ is the set

$$\mathrm{Spec}(\hat{H}) = \{\lambda \in \mathbb{C}\,\text{s.t.}\ \hat{H} - \lambda\hat{I}\ \text{is non invertible}\}. \tag{3.1}$$

For a finite-dimensional Hilbert space this is identical to the set of all finitely many eigenvalues of $\hat{H}$.

Using basic result from linear algebra, the spectrum of a self-adjoint operator $\hat{M}$ is a set of real eigenvalues λ_n, each with a corresponding eigenstate $|n\rangle$. Eigenstates corresponding to different eigenvalues are automatically orthogonal. If there is degeneracy then for each eigenspace of dimension greater than one we can always choose a basis of orthogonal eigenstates (Apply Grahm-Schmidt procedure eigenspace by eigenspace). Of course, we can always normalise the eigenstates, then we will have N orthonormal eigenstates giving us an orthonormal basis for $\mathcal{H}$. This also give us the *spectral representation* of $\hat{M}$ (corresponding to diagonalisation of the matrix M)

The spectral representation of an operator expresses it in terms of its eigenvectors and eigenvalues.

$$\hat{M} = \sum_n \lambda_n |n\rangle \langle n| \, .$$

Note that for the identity operator, the only eigenvalue is 1 with degeneracy N, so we can choose any orthonormal basis of $\mathcal{H}$ and

$$\hat{I} = \sum_n |n\rangle \langle n| \, .$$

This is a very useful expression. We can often use it in calculations by "inserting the identity as a complete sum of states."

Now when a measurement of M is made on a state

$$|\psi\rangle = \sum_n c_n |n\rangle \ ,$$

we will get the result λ_n with probability $p_n = \left|\left\langle n \middle| \hat{\psi} \right\rangle\right|^2 = |c_m|^2$ which is just the magnitude squared of the coefficient of $|n\rangle$ if we write ψ in the basis $\{|n\rangle\}$. After the measurement, if the result is λ_n, the state will then have definite value of M, λ_n, so measuring M again will give the same result. Therefore the state is no longer $|\psi\rangle$ but is $|n\rangle$. Note that this "collapse of the wavefunction" is not a unitary process, and is not reversible.

One way to describe this measurement process is in terms of the set of projection operators $\hat{P}_n = |n\rangle \langle n|$ formed from the eigenstates of $\hat{M}$. Then the probability of result λ_n is $p_n = \left\langle \psi \middle| \hat{P}_n \middle| \psi \right\rangle$ and the resulting state is $\frac{1}{\sqrt{p_n}} \hat{P}_n |\psi\rangle$ which is the state $|n\rangle$ up to an irrelevant overall phase.

Remember that a projector $\hat{P}$ is a linear operator such that $\hat{P}^\dagger = \hat{P}$ and $\hat{P}^2 = \hat{P}$ and check that indeed $\hat{P}_n = |n\rangle \langle n|$ has all these properties.

Projection operators can be used to express the measurement process.

The above discussion of measurement assume the spectrum of $\hat{M}$ is not degenerate. If we have degeneracy then we can generalise the definition of the projection operators. Consider an eigenvalue λ. We define the projection operator to be a sum over the eigenstates with that eigenvalue, i.e.

$$\hat{P}_\lambda = \sum_{n:\lambda_n=\lambda} |n\rangle \langle n| \, .$$

Then we still have the result that the probability is $p_\lambda = \left\langle \psi \middle| \hat{P}_\lambda \middle| \psi \right\rangle$ and the resulting state is $\frac{1}{\sqrt{p_\lambda}} \hat{P}_\lambda |\psi\rangle$.

An important point to note is that a state can only have definite values for two observables, say A and B, if it is a simultaneous eigenstate of $\hat{A}$ and $\hat{B}$. This is not possible for two generic operators. However, if $\left[\hat{A}, \hat{B}\right] = 0$ then we can always find simultaneous eigenstates. In this case we say that the observables A and B are *compatible*. If the observables are not compatible then measuring A, then B, then A again will not necessarily give the same result for the two measurements of A. That is because the state after the first measurement of A is not an eigenstate of $\hat{B}$, and so a measurement of B will change the state (to some eigenstate of $\hat{B}$.) This will not be an eigenstate of A, so the result of the second measurement of A cannot be determined with certainty.

Example:

Let $\mathcal{H} = \text{span}\{|-2\rangle, |-1\rangle, |1\rangle, |2\rangle\}$ be a four dimensional Hilbert space with orthonormal basis vectors given by the eigenvectors of an hermitian operator $\hat{A}$ as

$$\begin{aligned} \hat{A}\,|-2\rangle &= -2\,|-2\rangle \ , \qquad & \hat{A}\,|-1\rangle &= -1\,|-1\rangle \ , \\ \hat{A}\,|1\rangle &= 1\,|1\rangle \ , \qquad & \hat{A}\,|2\rangle &= 2\,|2\rangle \ , \end{aligned}$$

giving the spectral decomposition

$$\begin{aligned} \hat{A} &= -2\,|-2\rangle\langle -2| - 1\,|-1\rangle\langle -1| + 1\,|1\rangle\langle 1| + 2\,|2\rangle\langle 2| \\ &= (-2)\hat{P}_{-2} + (-1)\hat{P}_{-1} + (+1)\hat{P}_{+1} + (+2)\hat{P}_{+2} \\ &= \sum_{\lambda \in \text{Spec}(\hat{A})} \lambda \hat{P}_{\lambda} \ . \end{aligned}$$

in terms of the projectors $\hat{P}_{\lambda} = |\lambda\rangle\langle\lambda|$.

If we prepare a state $|\psi\rangle \in \mathcal{H}$ and measure the observable $\hat{A}$ we can only find one of the values $\{-2, -1, 1, 2\}$.

Suppose we prepared the state

$$|\psi\rangle = 2\,|-2\rangle + (1+i)\,|-1\rangle + 3i\,|1\rangle$$

we know that if we were to measure $\hat{A}$ on $|\psi\rangle$ we will never find the outcome $+2$ since the coefficient of $|2\rangle$ in the expansion for $|\psi\rangle$ vanishes.

To compute the probabilities of measuring $\{-2, -1, 1, 2\}$ we have to normalise the state, i.e. we need to impose $\langle\psi\,|\psi\rangle = 1$. We compute

$$\begin{aligned} \langle\psi\,|\psi\rangle = &\Big[2\,\langle -2| + (1-i)\,\langle -1| + (-3i)\,\langle 1|\Big] \\ &\times \Big[2\,|-2\rangle + (1+i)\,|-1\rangle + 3i\,|1\rangle\Big] = 4 + 2 + 9 = 15 \qquad (3.2) \end{aligned}$$

and use this result to define the normalised state

$$\left|\tilde{\psi}\right\rangle = \frac{|\psi\rangle}{\sqrt{15}} = \frac{2}{\sqrt{15}}\,|-2\rangle + \frac{(1+i)}{\sqrt{15}}\,|-1\rangle + \frac{3i}{\sqrt{15}}\,|1\rangle \ .$$

The probability of measuring $\hat{A}$ and finding outcome -2 is then the modulus square of the coefficient in front of $|-2\rangle$, i.e. $p_{-2} = \frac{4}{15}$, similarly $p_{-1} = \frac{2}{15}$, $p_{+1} = \frac{9}{15}$ and of course $p_{+2} = 0$. The total probability is 1 as it should since $\left\langle \tilde{\psi} \middle| \tilde{\psi} \right\rangle = 1$. Using the projector $\hat{P}_{-2} = |-2\rangle\langle -2|$ we have $p_{-2} = \left\langle \tilde{\psi} \right| \hat{P}_{-2} \left| \tilde{\psi} \right\rangle$.

If we prepare many copies of the same state $|\psi\rangle$, measure $\hat{A}$ and then average, we

> find the expectation value
>
> $$\begin{aligned}\langle A\rangle_\psi &= \frac{\langle\psi|\,\hat{A}\,|\psi\rangle}{\langle\psi\,|\psi\rangle} = \left\langle\tilde{\psi}\middle|\,\hat{A}\,\middle|\tilde{\psi}\right\rangle \\ &= \frac{4}{15}\,\langle -2|\,\hat{A}\,|{-2}\rangle + \frac{2}{15}\,\langle -1|\,\hat{A}\,|{-1}\rangle + \frac{9}{15}\,\langle 1|\,\hat{A}\,|1\rangle \\ &= p_{-2}(-2) + p_{-1}(-1) + p_{+1}(+1) + p_{+2}(+2) = -\frac{1}{15}\,.\end{aligned}$$

3.2 Density matrices

LECTURE 5

The above sections give the standard Dirac notation description of QM. The states previously described are what we will now call *pure states*. This means that the states are definite, i.e. we assume that (at least in principle) we know what the state of the system is. Any uncertainties in predictions are due to the nature of QM. However, we can also consider *mixed states* which arise when we do not know with certainty the state of a system. Here we assume that we have some probabilistic knowledge, such as the system is in state $|\psi\rangle$ with probability p, and in state $|\phi\rangle$ with probability $1-p$. This type of uncertainty is 'classical uncertainty' in the sense that it just describes our lack of knowledge about a system. Indeed, whether the state is pure or mixed may be a matter of perspective since one person may have more knowledge about the system than other (we will see this later when we discuss the reduced density matrix for a bipartite system).

> Pure states are fully determined, mixed states arise when we do not know the state of a system. A sum of basis states is still a pure state.

For a pure state $|\psi\rangle$ we define the *density operator* or, as more commonly called, the *density matrix* to be

$$\hat{\rho} = |\psi\rangle\,\langle\psi|\,.$$

Note that when our Hilbert space is n-dimensional if we think of ket vectors $|\psi\rangle$ as n components column vectors $\mathbf{z}$ and bra vectors $\langle\phi|$ as n components row vectors $\mathbf{w}^\dagger$, then an operator of the form $|\psi\rangle\,\langle\phi|$ can be thought of as $\mathbf{z}\mathbf{w}^\dagger$, hence a $n\times 1$ matrix times a $1\times n$ matrix, i.e. a $n\times n$ matrix, while the inner product $\mathbf{w}^\dagger\mathbf{z}$ as a $1\times n$ matrix times a $n\times 1$ matrix resulting in a 1×1 matrix, i.e. a complex number.

For pure states there is a one-to-one mapping between the density matrix and the state, so we can work with one or the other. For example we have the following correspondence:

$$\begin{aligned}\hat{M}\,|\psi\rangle = \lambda\,|\psi\rangle \quad &\longleftrightarrow \quad \hat{M}\hat{\rho} = \lambda\hat{\rho} \\ |\psi\rangle \to \hat{U}\,|\psi\rangle \quad &\longleftrightarrow \quad \hat{\rho} \to \hat{U}\hat{\rho}\hat{U}^\dagger\end{aligned}$$

Inner products of states arise when multiplying operators or when taking traces. In particular, if we label the orthonormal basis states $|n\rangle$ for some range of integers n, we define the trace of $\hat{A}$ to be:

$$\mathrm{Tr}(\hat{A}) = \sum_n \left\langle n\middle|\hat{A}\middle|n\right\rangle\,,$$

you can think of $\left\langle m\middle|\hat{A}\middle|n\right\rangle$ as the m^{th} row, n^{th} column entry of the matrix representation of $\hat{A}$ operator in the standard basis, hence the trace just defined corresponds indeed to the sum of the diagonal entries. Note that

> The trace of the density matrix $\hat{\rho}$ is equal to one, $\mathrm{Tr}(\hat{\rho}) = 1$.

$$\begin{aligned}\mathrm{Tr}(\hat{\rho}) &= \sum_n \langle n\,|\hat{\rho}|\,n\rangle = \sum_n \langle n\,|\psi\rangle\,\langle\psi\,|n\rangle \\ &= \sum_n \langle\psi\,|n\rangle\,\langle n\,|\psi\rangle = \left\langle\psi\,\middle|\hat{I}\middle|\,\psi\right\rangle = 1\,,\end{aligned} \tag{3.3}$$

where in the last two steps we used the spectral representation of the identity operator and the fact that $|\psi\rangle$ is normalised. Similar manipulations show that in general $\mathrm{Tr}(|\phi\rangle\,\langle\psi|) = \langle\psi\,|\phi\rangle$. Also note that for a pure state $\mathrm{Tr}(\hat{\rho}^2) = 1$ since $\hat{\rho}$ is a projector and we know that for projectors we have $\hat{\rho}^2 = \hat{\rho}$.

Mixed states describe situations where there is uncertainty about the state of the system due to lack of knowledge, i.e. this is the usual 'classical' uncertainty we have if we don't know everything about the system. We can describe mixed states in terms of an ensemble of pure states, each with a given probability of being the state of the system, e.g. $\{(p_i, |i\rangle)\}$ with $|i\rangle$ not necessarily orthogonal but chosen with unit norm (if not just normalised them one by one). The density matrix is just the linear combination of the density matrices for each of the pure states, weighted by the probability, i.e.

$$\hat{\rho} = \sum_i p_i\,|i\rangle\,\langle i|\,.$$

The density operator (or density matrix) can be used to describe the state of a system, for both pure and mixed states.

Note that there is no requirement for the state $|i\rangle$ to be orthogonal (although we assume they are normalised) and also such a mixed state density matrix does not correspond to a unique ensemble. There will be in general more than one ensemble $\{(p_i, |i\rangle)\}$ giving rise to the same density matrix for the same mixed state.

Of course, the probabilities p_i cannot be negative and must sum to 1. We can also generalise the definition of the mixed state density matrix to allow ensembles including mixed states.I.e. we can have $\hat{\rho} = \sum_i p_i\hat{\rho}_i$ where the $\hat{\rho}_i$ are mixed and/or pure state density matrices.

Such ensembles can only give a pure state in the trivial case where there is only one pure state, which must then have probability 1. However, given a density matrix it is often not immediately obvious whether it describes a pure or a mixed state. A test for this (see later discussion) is to calculate $\mathrm{Tr}(\hat{\rho}^2)$ which will be 1 for a pure state and less than 1 for a mixed state.

By construction, density matrices are

- normalised such that Tr $\hat{\rho} = 1$;
- Hermitian $\hat{\rho}^\dagger = \hat{\rho}$;
- *positive operators*, meaning that for any state $|\psi\rangle$, $\langle\psi\,|\hat{\rho}|\,\psi\rangle \geq 0$ (Note this can be equal to zero even for $|\psi\rangle \neq 0$. In matrix language this is called *semi*-positive-definite).

If we measure then the results for pure states in Dirac notation generalise to all pure or mixed density matrices as:

- The value of the result is λ with probability $p_\lambda = \mathrm{Tr}(\hat{P}_\lambda\,\hat{\rho}) = \mathrm{Tr}(\hat{P}_\lambda\,\hat{\rho}\,\hat{P}_\lambda)$.
- The density matrix after measuring M to be λ is

$$\hat{\rho} \to \frac{1}{p_\lambda}\,\hat{P}_\lambda\,\hat{\rho}\,\hat{P}_\lambda = \frac{1}{\mathrm{Tr}(\hat{P}_\lambda\,\hat{\rho}\,\hat{P}_\lambda)}\,\hat{P}_\lambda\,\hat{\rho}\,\hat{P}_\lambda\,.$$

Example:

Given a two dimensional Hilbert space $\mathcal{H} = \text{span}\{|0\rangle, |1\rangle\}$ decide whether the matrix $\rho = \frac{1}{9}\begin{pmatrix} 5 & 2-4i \\ 2+4i & 4 \end{pmatrix}$ is

1) a density matrix for a pure state,

2) a density matrix for a mixed state,

3) not a density matrix,

when we represented the basis vector using the standard basis.

First of all to be a density matrix the matrix ρ needs to be with $\text{Tr}\rho = 1$, hermitian $\rho^\dagger = \rho$ and semi-positive definite, i.e. $\mathbf{z}^\dagger \rho\, \mathbf{z} \geq 0$ for all $\mathbf{z} \in \mathbb{C}^2$.

It is simple to check that ρ is indeed hermitian and with trace equal to 1. Instead of checking that ρ is semi-positive definite let us see how a density matrix for a pure state $|\psi\rangle = a\,|0\rangle + b\,|1\rangle$ looks like. Let us assume the state is normalised so $|a|^2 + |b|^2 = 1$ and pass to vector representation $|\psi\rangle \to \begin{pmatrix} a \\ b \end{pmatrix}$ then the matrix associate to its density operator is

$$\hat{\rho_\psi} = |\psi\rangle\langle\psi| \to \rho_\psi = \begin{pmatrix} a \\ b \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}^\dagger = \begin{pmatrix} |a|^2 & ab^* \\ ba^* & |b|^2 \end{pmatrix}.$$

It is simple to see that if we chose $b = \frac{2}{3}$ and $a = \frac{1-2i}{3}$ we obtain precisely the matrix ρ under question. Note that this is not the only possibility! We can multiply $|\psi\rangle = \frac{1-2i}{3}|0\rangle + \frac{2}{3}|1\rangle$ by any phase $e^{i\alpha}$ without changing its density matrix.

3.3 Pure versus mixed states

In the example above we were able to find explicitly the pure state whose density operator was the matrix provided, however we would like to know whether a certain density operator given comes from a pure state or a mixed one.To this end we have the following theorem.

Theorem: Let $\hat{\rho}$ be a density operator on a Hilbert space $\mathcal{H}$, i.e. $\mathrm{Tr}\,\hat{\rho} = 1$, $\hat{\rho}^\dagger = \hat{\rho}$ and $\hat{\rho}$ positive operator. The density operator $\hat{\rho}$ corresponds to a pure state if and only if $\mathrm{Tr}\,\hat{\rho}^2 = 1$.

Proof:

($\Rightarrow$) Let us assume that $\hat{\rho} = |\psi\rangle\langle\psi|$ is the density matrix associated to a pure state $|\psi\rangle \in \mathcal{H}$. It is simple to compute $\hat{\rho}^2 = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = \hat{\rho}$ since the state is normalised, hence $\mathrm{Tr}\,\hat{\rho}^2 = \mathrm{Tr}\hat{\rho} = 1$.

($\Leftarrow$) Conversely let us suppose that $\hat{\rho}$ is the density operator corresponding to the ensemble $\{p_i, |\psi_i\rangle\}$, i.e. $\hat{\rho} = \sum_i p_i \hat{\rho}_i = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. We want to compute $\mathrm{Tr}\hat{\rho}^2$:

$$\begin{aligned}
\mathrm{Tr}\hat{\rho}^2 &= \sum_n \langle n|\hat{\rho}^2|n\rangle = \sum_{n,i,j} p_i p_j \langle n|\psi_i\rangle\langle\psi_i|\psi_j\rangle\langle\psi_j|n\rangle \\
&= \sum_{ij} p_i p_j \langle\psi_i|\psi_j\rangle \left(\sum_n \langle\psi_j|n\rangle\langle n|\psi_i\rangle\right) \\
&= \sum_{ij} p_i p_j \langle\psi_i|\psi_j\rangle\langle\psi_j|\hat{I}|\psi_i\rangle \\
&= \sum_{ij} p_i p_j \langle\psi_i|\psi_j\rangle\langle\psi_j|\psi_i\rangle \\
&= \sum_{ij} p_i p_j |\langle\psi_i|\psi_j\rangle|^2 \leq \sum_{ij} p_i p_j \leq 1\,.
\end{aligned}$$

In the second line we used the spectral decomposition of the identity operator $\hat{I} = \sum_n |n\rangle\langle n|$, while in the third line we made use of the complex Cauchy-Schwarz inequality

$$|\langle\psi_i|\psi_j\rangle|^2 \leq \langle\psi_i|\psi_i\rangle\langle\psi_j|\psi_j\rangle \leq 1$$

since the states $|\psi_i\rangle$ are normalised. Finally in the last step we used the fact that the p_i are probabilities and $\sum_i p_i = 1$.

For a pure state, the density matrix has $\mathrm{Tr}(\hat{\rho}^2) = 1$. For a mixed state, we have instead $\mathrm{Tr}(\hat{\rho}^2) < 1$.

We also know that the equality in the Cauchy-Schwarz inequality holds if and only if the vectors $|\psi_i\rangle$ and $|\psi_j\rangle$ are collinear, i.e. $|\psi_i\rangle = a|\psi_j\rangle$ for some complex number $a \in \mathbb{C}$ that can only be a phase $a = e^{i\alpha}$ since all the vectors must have length one.

Hence we have that $\mathrm{Tr}\,\hat{\rho}^2 \leq 1$ with equality if and only if all vectors are collinear with one another, i.e. they are all a multiple of say the first one $|\psi_i\rangle = e^{i\alpha_i}|\psi_1\rangle$ but this means that the density matrix

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i| = \sum_i p_i |\psi_1\rangle\langle\psi_1| = |\psi_1\rangle\langle\psi_1|\,,$$

hence Tr $\hat{\rho}^2 = 1$ and $\hat{\rho}$ is a pure state.

We then have a complete characterization of pure vs mixed states! We just need to compute Tr $\hat{\rho}^2$ if this number is less than one we know that we have a mixed state, if we find one we know the state is pure.

We will shortly give a geometric characterization for pure and mixed state in the simplest case of a two dimension Hilbert space, i.e. what we call a qubit.

> **Example:**
>
> Suppose we have a three dimensional Hilbert space with orthonormal basis $\mathcal{H} = \text{span}\{|1\rangle, |3\rangle, |5\rangle\}$. Compute the density matrix associated to the ensemble $\{(2/3, |\psi_1\rangle), (1/3, |\psi_2\rangle)\}$ where $|\psi_1\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |3\rangle)$ and $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|3\rangle + i\,|5\rangle)$.
>
> First of all we notice that the state $|\psi_1\rangle, |\psi_2\rangle$ are normalised, had they not we would have had to normalise them before proceeding. The density operator associated with this mixed state is then
>
> $$\begin{aligned}\hat{\rho} &= \frac{2}{3}\,|\psi_1\rangle\langle\psi_1| + \frac{1}{3}\,|\psi_2\rangle\langle\psi_2| \\ &= \frac{2}{6}(|1\rangle - |3\rangle)(\langle 1| - \langle 3|) + \frac{1}{6}(|3\rangle + i\,|5\rangle)(\langle 3| - i\,\langle 5|) \\ &= \frac{2}{6}\,|1\rangle\langle 1| - \frac{2}{6}\,|1\rangle\langle 3| - \frac{2}{6}\,|3\rangle\langle 1| \\ &\quad + \frac{1}{2}\,|3\rangle\langle 3| + \frac{i}{6}\,|3\rangle\langle 5| - \frac{i}{6}\,|5\rangle\langle 3| + \frac{1}{6}\,|5\rangle\langle 5|\,.\end{aligned}$$
>
> If we represent the three basis vectors using the standard basis we can write the density operator as the 3×3 matrix
>
> $$\begin{aligned}\rho &= \frac{2}{3}\begin{pmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\\ 0\end{pmatrix}\begin{pmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\\ 0\end{pmatrix}^{\dagger} + \frac{1}{3}\begin{pmatrix}0\\ 1/\sqrt{2}\\ i/\sqrt{2}\end{pmatrix}\begin{pmatrix}0\\ 1/\sqrt{2}\\ i/\sqrt{2}\end{pmatrix}^{\dagger} = \\ &= \begin{pmatrix}\frac{1}{3} & -\frac{1}{3} & 0\\ -\frac{1}{3} & \frac{1}{2} & \frac{i}{6}\\ 0 & -\frac{i}{6} & \frac{1}{6}\end{pmatrix}.\end{aligned}$$
>
> It is simple to check now that the trace of this matrix is of course one, while $\text{Tr}\rho^2 = \frac{2}{3} < 1$ since the state is a mixed state.
>
> Finally if we have the observable $\hat{A}$ with spectrum $\{1, |1\rangle; 3, |3\rangle; 5, |5\rangle\}$, we can easily compute the expectation value on this state
>
> $$\text{Tr}\left[\hat{\rho}\hat{A}\right] = \text{Tr}\left[\rho\begin{pmatrix}1 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 5\end{pmatrix}\right] = \frac{8}{3}\,,$$
>
> which you can also check using the abstract operator formalism.

> **Example:**
>
> Let $\mathcal{H} = \text{span}\{|1\rangle, ..., |6\rangle\}$ be a six dimensional Hilbert space with orthonormal basis $|i\rangle$ given by the eigenvectors with eigenvalues $\{1, 2, 3, 4, 5, 6\}$ for the her-

mitian operator $\hat{A}$. Consider the normalised mixed state given by the ensemble $\{(\frac{1}{6}, |1\rangle), (\frac{1}{6}, |2\rangle), (\frac{1}{6}, |3\rangle), (\frac{1}{6}, |4\rangle), (\frac{1}{6}, |5\rangle), (\frac{1}{6}, |6\rangle)\}$

$$\hat{\rho} = \sum_{i=1}^{6} \frac{1}{6} |i\rangle \langle i| = \frac{1}{6}\hat{I} \,, \qquad \rho = \frac{\mathbb{I}_6}{6} \,,$$

where we used the standard basis to represent the basis vectors and obtain the matrix representation for $\hat{\rho}$ given by ρ.

This state is in a certain sense (that we will quantify later on) the most mixed, it is an equally probable ensemble of the six basis vectors. Its trace is clearly one while $\text{Tr}(\hat{\rho}^2) = \text{Tr}(\rho^2) = \frac{1}{6} < 1$.

The expectation value of the observable $\hat{A}$, with spectrum precisely $\{1, |1\rangle \,; \ldots ; 6, |6\rangle\}$, on this state is given by

$$\langle A \rangle = \text{Tr}(\hat{\rho}\hat{A}) = \sum_{i=1}^{6} \frac{1}{6} \times i = \frac{7}{2} \,,$$

if you want the state $\hat{\rho}$ is the most “classically” uncertain of all the states, it is exactly the same ensemble of a six-faced die for which the average outcome is precisely $7/2$.

4 Qubits and the Bloch sphere

LECTURE 7

4.1 Qubits

In this chapter we want to understand the smallest dimensional (yet extremely interesting) quantum system. Because of the equivalence relation $|\psi\rangle \sim c\,|\psi\rangle$, a one-dimensional Hilbert space is trivial since it only describes a single state. Therefore the smallest non-trivial system has dimension two, and in QI we refer to such a system as a *qubit*. A standard orthonormal basis is labelled $\{|0\rangle\,, |1\rangle\}$ (sometimes you will also find it written as $\{|\uparrow\rangle\,, |\downarrow\rangle\}$ denoting the two states of a spin 1/2 particle with spin "up" or "down") so we see the analogy with a classical bit which can take either value 0 or 1 hence we usually call the basis $\{|0\rangle\,, |1\rangle\}$ the *computational basis*. The difference for the qubit is that the state can be a linear combination of $|0\rangle$ and $|1\rangle$. The qubit is important since many general features can be understand in this simple Hilbert space. Also, larger systems can often be described a being built from several qubits – this is used in quantum information and especially in quantum computation.

So let us try and characterise the most general pure state of the qubit Hilbert space $\mathcal{H}_q = \text{span}\{|0\rangle\,, |1\rangle\}$. Now if we have a qubit system the most general pure state can be written as a linear combination of the two basis vectors

$$|\psi\rangle = a\,|0\rangle + b\,|1\rangle\ ,$$

with $a, b \in \mathbb{C}$ and $|a|^2 + |b|^2 = 1$ so that the state is already normalised $\langle\psi\,|\psi\rangle = 1$.

Up to an irrelevant overall phase we can assume that $a \in \mathbb{R}$, hence any normalised pure state can be written in the form

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

The most general qubit state is labelled by two complex numbers, subject to a norm constraint and the fact that the overall phase is irrelevant. This leaves two real parameters, which parametrise a sphere.

where $0 \le \theta \le \pi$ and $0 \le \phi < 2\pi$. Note that the normalisation condition $|a|^2 + |b|^2 = 1$ now becomes $\cos\left(\frac{\theta}{2}\right)^2 + |e^{i\phi}\sin\left(\frac{\theta}{2}\right)|^2 = 1$.

This new coordinates (θ, ϕ) are just a convenient parametrisation of $|a|^2 + |b|^2 = 1$ with $a \in \mathbb{R}$ in terms of angles which correspond to the angles in spherical polar coordinates. Indeed the equation $|a|^2 + |b|^2 = 1$ with $a \in \mathbb{R}$ can be rewritten as $a^2 + (\text{Re}\, b)^2 + (\text{Im}\, b)^2 = 1$ which is indeed the equation describing a 2-dimensional sphere, what we call S^2. Therefore we have a one-to-one mapping between pure qubits states and points on a sphere (say of radius 1.) This is called the *Bloch Sphere*.

Now any point on the Bloch sphere can also be labelled by its position vector, called the *Bloch vector* $\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ which has $|\mathbf{r}| = \sqrt{x^2 + y^2 + z^2} = 1$. The dictionary between Cartesian coordinates and polar coordinates is very simple

$$x = \sin(\theta)\cos(\phi)\,, \quad y = \sin(\theta)\sin(\phi)\,, \quad z = \cos(\theta)\,,$$

θ is the polar angle measured from the z-axis, while ϕ is the azimuthal angle measured around the z-axis as in Figure 4.1.

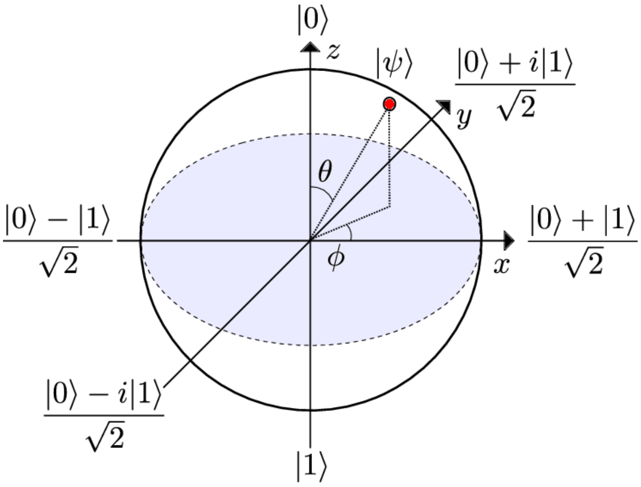

$|0\rangle$
z
$|\psi\rangle$
$\frac{|0\rangle + i|1\rangle}{\sqrt{2}}$
y
θ
$\frac{|0\rangle - |1\rangle}{\sqrt{2}}$
$\frac{|0\rangle + |1\rangle}{\sqrt{2}}$
ϕ
x
$\frac{|0\rangle - i|1\rangle}{\sqrt{2}}$
$|1\rangle$

Figure 4.1: Graphic representation of the Bloch sphere. The states $|0\rangle\,, |1\rangle$ are on the z-axis, the state $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ are on the x-axis, and the states $|L\rangle = (|0\rangle + i\,|1\rangle)/\sqrt{2}\,, |R\rangle = (|0\rangle - i\,|1\rangle)/\sqrt{2}$ are on the y-axis.

There are six special states on the Bloch sphere whose cartesian coordinates are

There are six special states in the qubit system, which are eigenvectors of three special operators (to be discussed later).

simply $\begin{pmatrix}\pm1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\\pm1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\\pm1\end{pmatrix}$ given by

$$
\begin{aligned}
\mathbf{r} &= \begin{pmatrix}1\\0\\0\end{pmatrix}\,, \qquad (\theta,\phi) = (\pi/2, 0)\,,\\
&\quad \rightarrow \quad |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \rightarrow \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}\,,\\
\mathbf{r} &= \begin{pmatrix}-1\\0\\0\end{pmatrix}\,, \qquad (\theta,\phi) = (\pi/2, \pi)\,,\\
&\quad \rightarrow \quad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \rightarrow \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}\,,\\
\mathbf{r} &= \begin{pmatrix}0\\1\\0\end{pmatrix}\,, \qquad (\theta,\phi) = (\pi/2, \pi/2)\,,\\
&\quad \rightarrow \quad |L\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle) \rightarrow \frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix}\,,\\
\mathbf{r} &= \begin{pmatrix}0\\-1\\0\end{pmatrix}\,, \qquad (\theta,\phi) = (\pi/2, 3\pi/2)\,,\\
&\quad \rightarrow \quad |R\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i\,|1\rangle) \rightarrow \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}\,,\\
\mathbf{r} & \begin{pmatrix}0\\0\\1\end{pmatrix}\,, \qquad (\theta,\phi) = (0, \cdot)\,,\\
&\quad \rightarrow \quad |0\rangle \rightarrow \begin{pmatrix}1\\0\end{pmatrix}\,,\\
\mathbf{r} &= \begin{pmatrix}0\\0\\-1\end{pmatrix}\,, \qquad (\theta,\phi) = (\pi, \cdot)\,,\\
&\quad \rightarrow \quad |1\rangle \rightarrow \begin{pmatrix}0\\1\end{pmatrix}\,.
\end{aligned}
$$

These states are special as they lie at the intersection of the Bloch sphere with the cardinal axis and we will see later on they will each have well defined measurements for three special observables for the qubit system.

> **Exercise:**
>
> Show that besides the computational basis $\{|0\rangle\,, |1\rangle\}$, the qubit Hilbert space $\mathcal{H}_q$ can be described in terms of the orthonormal basis $\{|+\rangle\,, |-\rangle\}$ or by the orthonormal basis $\{|L\rangle\,, |R\rangle\}$.

4.2 Inside the Bloch sphere

We want to rewrite now the density matrix for a general pure state written using polar angles to cartesian coordinates. To this end we start with a generic pure state $|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$ and we first rewrite it using as above the standard basis representation for the two basis ket vector $|0\rangle \to \begin{pmatrix}1\\0\end{pmatrix}$, $|1\rangle \to \begin{pmatrix}0\\1\end{pmatrix}$ hence we have

$$
\begin{aligned}
|\psi\rangle &\to \begin{pmatrix}\cos\left(\frac{\theta}{2}\right)\\ e^{i\phi}\sin\left(\frac{\theta}{2}\right)\end{pmatrix}\,,\\
\hat{\rho} = |\psi\rangle\langle\psi| \to \rho &= \begin{pmatrix}\cos\left(\frac{\theta}{2}\right)\\ e^{i\phi}\sin\left(\frac{\theta}{2}\right)\end{pmatrix}\begin{pmatrix}\cos\left(\frac{\theta}{2}\right) & e^{-i\phi}\sin\left(\frac{\theta}{2}\right)\end{pmatrix}\\
\rho &= \frac{1}{2}\begin{pmatrix}1+\cos(\theta) & e^{-i\phi}\sin(\theta)\\ e^{+i\phi}\sin(\theta) & 1-\cos(\theta)\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1+z & x-iy\\ x+iy & 1-z\end{pmatrix}\,,
\end{aligned}
$$

where in the last line we used some simple trigonometric identities and finally the dictionary spherical ↔ cartesian coords.

It can be seen quite easily now that the density matrix can be written in terms of the Bloch vector $\mathbf{r} = \begin{pmatrix}x\\y\\z\end{pmatrix}$ as $\rho = \frac{1}{2}\left(\mathbb{I}_2 + \mathbf{r}\cdot\sigma\right)$, where we introduce the three matrices σ_i called the Pauli σ-matrices

$$
\sigma_1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\,,\quad \sigma_2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}\,,\quad \sigma_3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\,,
$$

so $\mathbf{r}\cdot\sigma = r_1\sigma_1 + r_2\sigma_2 + r_3\sigma_3$.

A state with given Bloch vector can be mapped to a density matrix using the Pauli matrices.

As the density matrix is linear in the Bloch vector, mixed states also have Bloch vector given in terms of a linear combination of the Bloch vectors of the states in the ensemble. Since the coefficients are probabilities, it is easy to see that the Bloch vector for a mixed state must be the position vector of a point inside the Bloch sphere.

Let us check this in detail. Suppose we have the ensemble $\{(p_i, \rho_i)\}$ for some probabilities p_i and pure states ρ_i defined by vectors $\mathbf{r}_i$ on the Bloch sphere, i.e. $|\mathbf{r}_i|^2 = 1$. The density matrix for this mixed state is given by

$$
\begin{aligned}
\rho &= \sum_i p_i\rho_i = \sum_i p_i \frac{(\mathbb{I}_2 + \mathbf{r}_i\cdot\sigma)}{2}\\
&= \frac{1}{2}\left(\mathbb{I}_2 + \mathbf{r}\cdot\sigma\right)\,,
\end{aligned}
\tag{4.1}
$$

where we defined the vector $\mathbf{r} = \sum_i p_i\mathbf{r}_i$. As said above, since the density matrix is linear in the Bloch vector, mixed states as well can be represented in terms of a Bloch vector $\mathbf{r}$. Let us check now where this vector lies:

$$
\begin{aligned}
|\mathbf{r}|^2 &= |\sum_i p_i\mathbf{r}_i|^2 = \sum_{ij} p_i p_j\,\mathbf{r}_i\cdot\mathbf{r}_j\\
&\leq \sum_{ij} p_i p_j|\mathbf{r}_i||\mathbf{r}_j| \leq \sum_{ij} p_i p_j \leq 1\,,
\end{aligned}
\tag{4.2}
$$

where we used the real Cauchy Schwartz inequality. Note that equality holds if and only if all $\mathbf{r}_i$ are collinear hence all the density matrices ρ_i of the ensemble are given by the same density matrix, hence we do not really have a mixed state but rather a pure one.

For genuine mixed states the inequality is strict and so mixed states are defined by point inside the Bloch sphere, i.e. $|\mathbf{r}| < 1$.

The Bloch sphere gives us a nice geometrical interpretation for the full set of states of the qubit system: points on the sphere define pure states, points inside the sphere define mixed states.

Remember the previous test to distinguish whether a density matrix corresponds to a pure or a mixed state by considering whether $\mathrm{Tr}(\rho^2) = 1$ or < 1 respectively. The condition $\mathrm{Tr}(\rho^2) < 1$ for ρ to correspond to a mixed state has now for a qubit a geometric interpretation:

$$\mathrm{Tr}(\rho^2) = \frac{1}{2}\left(1 + |\mathbf{r}|^2\right) \leq 1\,,$$

mixed states are **inside** the Bloch sphere $|\mathbf{r}| < 1$, pure state are **on** the Bloch sphere $|\mathbf{r}| = 1$.

4.3 Time-evolution of a qubit

Unitary transformations of a qubit are represented as rotations of the Bloch sphere about the origin. This illustrates that unitary transformations cannot transform pure states to mixed states, but also we see that not all mixed states are related in this way. Indeed $\mathrm{Tr}(\rho^2) = (1 + |\mathbf{r}|^2)/2$ is invariant under unitary transformations, and is a measure of how mixed a state is, with $\mathrm{Tr}(\rho^2) = 1$ for pure states to $\mathrm{Tr}(\rho^2) = 1/2$ for the "most mixed" state corresponding to the origin, i.e. $\rho = \mathbb{I}_2/2$. Of course, measurements are not unitary transformations and they act as projectors hence they can transform any state to a pure state.

> **Exercise:**
>
> Show that the quantity $\mathrm{Tr}(\hat\rho^2)$ is invariant under time evolution.
>
> *Solution*: We simply need to remember that a density matrix $\hat\rho$ transforms under time evolution has $\hat\rho \to \hat U \hat\rho \hat U^\dagger$ hence
>
> $$\hat\rho^2 \to \hat U \hat\rho \hat U^\dagger \hat U \hat\rho \hat U^\dagger = \hat U \hat\rho^2 \hat U^\dagger$$
>
> using the fact that time evolution is a unitary operation $\hat U^\dagger \hat U = \hat I$. We then conclude that $\mathrm{Tr}(\hat U \hat\rho^2 \hat U^\dagger) = \mathrm{Tr}(\hat\rho^2)$ using cyclicity of the trace. Since $\mathrm{Tr}(\hat\rho^2)$ measures how mixed a state is (we will quantify better later on) it is then clear that time evolution cannot possibly change that.

> **Example:**
>
> Let $\mathbf{r}_1$ and $\mathbf{r}_2$ denote two distinct points on the Bloch sphere, i.e. $|\mathbf{r}_1| = |\mathbf{r}_1| = 1$ and $\mathbf{r}_1 \neq \mathbf{r}_2$, and consider the ensemble $\{(p, \mathbf{r}_1), (1-p, \mathbf{r}_2)\}$ with $0 \leq p \leq 1$. The density matrix corresponding to this ensemble is given by
>
> $$\rho = \rho_1 + \rho_2 = \frac{1}{2}(\mathbb{I}_2 + \mathbf{r}\cdot\sigma)\,,$$

> with $\mathbf{r} = p\,\mathbf{r}_1 + (1-p)\mathbf{r}_2$. Note that geometrically the Bloch vector $\mathbf{r}$ lies on the line between the points $\mathbf{r}_1$ and $\mathbf{r}_2$ and since $0 \leq p \leq 1$ the point $\mathbf{r}$ on this line will always fall within the Bloch sphere.
>
> The state ρ will be pure if and only if $p = 1$, in which case $\rho = \rho_1$, or $p = 0$, in which case $\rho = \rho_2$. Note in particular that mixing can never produce a state farther from the origin then the farthest initial state. Furthermore once we have chosen a mixed state, i.e. a point inside the Bloch sphere, we can find an infinite number of ways to write it as an ensemble of two pure states! We just need to consider any line passing through this point which will intersect the Bloch sphere in two points corresponding to the two pure states of which this mixed state is an ensemble of.
>
> In particular the "most" mixed of the qubit states is $\rho = \mathbb{I}_2/2$ which correspond of an ensemble of any two antipodal points, for example the North and South poles $|0\rangle\,, |1\rangle$ states (but any other two antipodal points on the sphere will produce the same), each one of them with probability 50%.

For a qubit system it is simple to define a "distance" between states given by the geometric distance of the relative positions using the Bloch sphere, i.e. $|\mathbf{r}_1 - \mathbf{r}_2|$ if the two states are described by the Bloch vectors $\mathbf{r}_1$, $\mathbf{r}_2$. This turns out to be equal to what is called the *trace distance* between the two states

$$D(\hat{\rho}_1, \hat{\rho}_2) = \frac{1}{2}\mathrm{Tr}|\hat{\rho}_1 - \hat{\rho}_2|\,,$$

where we need to give the definition of the operator $|\hat{A}|$.

The operator $|\hat{A}|$ is defined to be the positive operator

$$|\hat{A}| = \sqrt{\hat{A}^\dagger \hat{A}}\,.$$

Note (check) that the operator $\hat{A}^\dagger\hat{A}$ is both hermitian and positive hence its square root is well defined. For this, you just need to chose a basis that diagonalises $\hat{A}^\dagger\hat{A}$ with non-negative eigenvalues (because the operator is hermitian and positive); in this basis the operator $\sqrt{\hat{A}^\dagger\hat{A}}$ will then be diagonal with eigenvalues given by the square root of the eigenvalues of $\hat{A}^\dagger\hat{A}$.

In practical terms if the operator $\hat{A}$ is hermitian with eigenvalues a_i then $\hat{A}^\dagger\hat{A} = \hat{A}^2$ and we have that

$$\mathrm{Tr}|\hat{A}| = \sum_i |a_i|\,.$$

With this definition for $\mathrm{Tr}|\hat{A}|$ we can then easily compute the trace distance between two qubit states

$$\begin{aligned} D(\hat{\rho}_1, \hat{\rho}_2) &= \frac{1}{2}\mathrm{Tr}|\hat{\rho}_1 - \hat{\rho}_2| = \frac{1}{4}\mathrm{Tr}|(\mathbf{r}_1 - \mathbf{r}_2)\cdot\sigma| \\ &= \frac{1}{2}|\mathbf{r}_1 - \mathbf{r}_2|\,, \end{aligned} \tag{4.3}$$

and indeed as anticipated the trace distance for qubit states correspond precisely to the geometric distance in the Bloch sphere representation.

The "trace distance" between two states described by a density matrix corresponds to the geometric distance between the two states in the Bloch sphere representation.

We can rewrite the trace distance between two states as

$$\begin{aligned} D(\hat{\rho}_1,\hat{\rho}_2) &= \frac{1}{2}\mathrm{Tr}|\hat{\rho}_1-\hat{\rho}_2| \\ &= \frac{1}{2}\mathrm{Tr}\sqrt{(\hat{\rho}_1-\hat{\rho}_2)^\dagger(\hat{\rho}_1-\hat{\rho}_2)} \\ &= \frac{1}{2}\mathrm{Tr}\sqrt{(\hat{\rho}_1-\hat{\rho}_2)^2} = \frac{1}{2}\sum_i|\lambda_i|\,, \end{aligned} \tag{4.4}$$

where λ_i are the eigenvalues of the hermitian but not necessarily positive operator $\hat{\rho}_1-\hat{\rho}_2$. This defines a *metric* on the space of density matrices, i.e. something with the following properties

- it is non-negative $D(\hat{\rho}_1,\hat{\rho}_2)\geq 0$;
- it is symmetric $D(\hat{\rho}_1,\hat{\rho}_2)=D(\hat{\rho}_2,\hat{\rho}_1)$;
- it satisfies triangle inequality $D(\hat{\rho}_1,\hat{\rho}_3)\leq D(\hat{\rho}_1,\hat{\rho}_2)+D(\hat{\rho}_2,\hat{\rho}_3)$;
- it separates points $D(\hat{\rho}_1,\hat{\rho}_2)=0\Leftrightarrow\hat{\rho}_1=\hat{\rho}_2$.

Note that there are many notions of "distance" when discussing quantum mechanical states. The trace distance is a possible meaningful notion of distance between states in that it tells us how distinguishable with measurements two states are, i.e. the closer they are the less "distinguishable" they are by simply measuring observables. Another such notion is what is called *fidelity* although we will not cover this.

> **Example:**
>
> Compute the trace distance between the qubit states
>
> $$\begin{aligned} \hat{\rho}_1 &= \frac{1}{4}|0\rangle\langle 0|+\frac{3}{4}|1\rangle\langle 1| \to \rho_1 = \begin{pmatrix}\frac{1}{4} & 0\\ 0 & \frac{3}{4}\end{pmatrix}\,, \\ \hat{\rho}_2 &= \frac{2}{3}|0\rangle\langle 0|+\frac{1}{3}|1\rangle\langle 1| \to \rho_2 = \begin{pmatrix}\frac{2}{3} & 0\\ 0 & \frac{1}{3}\end{pmatrix}\,. \end{aligned}$$
>
> If we compute the matrix $\rho_1-\rho_2$ we obtain
>
> $$\begin{aligned} \rho_1-\rho_2 &= \begin{pmatrix}\frac{-5}{12} & 0\\ 0 & \frac{5}{12}\end{pmatrix} \\ \Rightarrow D(\hat{\rho}_1,\hat{\rho}_2) &= \frac{1}{2}\mathrm{Tr}|\rho_1-\rho_2| \\ &= \frac{1}{2}\left(\left|\frac{-5}{12}\right|+\left|\frac{5}{12}\right|\right)=\frac{5}{12}\,. \end{aligned} \tag{4.5}$$
>
> We could have also computed the two Bloch vectors associated to the two density matrices $\mathbf{r}_1=(0,0,-1/2)^T$ and $\mathbf{r}_2=(0,0,1/3)^T$ and in fact we have
>
> $$D(\hat{\rho}_1,\hat{\rho}_2)=\frac{1}{2}|\mathbf{r}_1-\mathbf{r}_2|=\frac{1}{2}\left|\begin{pmatrix}0\\0\\-\frac{5}{6}\end{pmatrix}\right|=\frac{5}{12}\,.$$

4.4 Pauli Matrices

We will summarise here some key properties of the so-called *Pauli matrices*, which were introduced above when we constructed the density matrix corresponding to a generic qubit state. The following can be checked easily.

$$\begin{aligned}
\sigma_1^\dagger = \sigma_1\,, \qquad \sigma_2^\dagger &= \sigma_2\,, \qquad \sigma_3^\dagger = \sigma_3\,,\\
\text{Tr}(\sigma_1) = \text{Tr}(\sigma_2) &= \text{Tr}(\sigma_3) = 0\,,\\
[\sigma_i,\sigma_j] &= \sigma_i\sigma_j - \sigma_j\sigma_i = 2i\,\epsilon_{ijk}\,\sigma_k\,,\\
\{\sigma_i,\sigma_j\} &= \sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}\mathbb{I}_2\,,\\
\sigma_i\sigma_j &= \delta_{ij}\mathbb{I}_2 + i\,\epsilon_{ijk}\sigma_k\,,
\end{aligned}$$

where δ_{ij} is the Kronecker delta and ϵ_{ijk} the Levi-Civita tensor.

So the Pauli matrices are Hermitian, traceless matrices, they actually form a basis for the vector space of 2×2 hermitian, traceless matrices (Ex. check that this is a vector space over the real numbers).

If we define the operators

$$\begin{aligned}
X &= \frac{1}{2}(\mathbb{I}_2 - \sigma_1) = \frac{1}{2}\begin{pmatrix} 1 & -1\\ -1 & 1\end{pmatrix}\,,\\
Y &= \frac{1}{2}(\mathbb{I}_2 - \sigma_2) = \frac{1}{2}\begin{pmatrix} 1 & i\\ -i & 1\end{pmatrix}\,,\\
Z &= \frac{1}{2}(\mathbb{I}_2 - \sigma_3) = \frac{1}{2}\begin{pmatrix} 0 & 0\\ 0 & 2\end{pmatrix}\,,
\end{aligned}$$

we can easily see that the six "special" states defined above are precisely the eigenvectors of these three operators with eigenvalues either 0 or 1, i.e.

$$\begin{aligned}
X\,|+\rangle &= 0\,|+\rangle\;, \qquad X\,|-\rangle = 1\,|-\rangle\;,\\
Y\,|L\rangle &= 0\,|L\rangle\;, \qquad Y\,|R\rangle = 1\,|R\rangle\;,\\
Z\,|0\rangle &= 0\,|0\rangle\;, \qquad Z\,|1\rangle = 1\,|1\rangle\;.
\end{aligned}$$

Finally another important property of the Pauli matrices is that their exponential are unitary matrices

$$\begin{aligned}
e^{i\alpha\sigma_1} &= \begin{pmatrix} \cos(\alpha) & i\sin(\alpha)\\ i\sin(\alpha) & \cos(\alpha)\end{pmatrix} = \cos(\alpha)\mathbb{I}_2 + i\sin(\alpha)\sigma_1\,,\\
e^{i\alpha\sigma_2} &= \begin{pmatrix} \cos(\alpha) & \sin(\alpha)\\ -\sin(\alpha) & \cos(\alpha)\end{pmatrix} = \cos(\alpha)\mathbb{I}_2 + i\sin(\alpha)\sigma_2\,,\\
e^{i\alpha\sigma_3} &= \begin{pmatrix} e^{i\alpha} & 0\\ 0 & e^{-i\alpha}\end{pmatrix} = \cos(\alpha)\mathbb{I}_2 + i\sin(\alpha)\sigma_3\,,
\end{aligned}$$

where $\alpha\in\mathbb{R}$.

> **Exercise:**
>
> Compute these exponentials and check that these are indeed unitary matrices.

More in general we can consider the unitary transformation

$$U_\alpha(\mathbf{n}) = e^{i\alpha\mathbf{n}\cdot\sigma} = \cos(\alpha)\mathbb{I}_2 + i\sin(\alpha)\mathbf{n}\cdot\sigma\,,$$

where again $\alpha \in \mathbb{R}$ and $\mathbf{n} \in \mathbb{R}^3$ is a 3-dimensional unit vector, i.e. $|\mathbf{n}|^2 = 1$.

This is a unitary transformation so it is a perfectly valid time evolution operator. If we now act on the density matrix $\rho = \frac{1}{2}\left(\mathbb{I}_2 + \mathbf{r} \cdot \sigma\right)$ of a state defined by the vector $\mathbf{r}$ on the Bloch sphere we can see that

$$U_\alpha(\mathbf{n}) \rho U_\alpha(\mathbf{n})^\dagger = \frac{1}{2}\left[\mathbb{I}_2 + (R_\alpha(\mathbf{n})\mathbf{r}) \cdot \sigma\right] ,$$

where $R_\alpha(\mathbf{n})$ is the 3×3 orthogonal matrix corresponding to a rotation of an angle 2α around the axis defined by the unit vector $\mathbf{n}$ acting on the three dimensional vector $\mathbf{r}$. This means that if we prepare our qubit say in the state $|0\rangle$ with Bloch vector $\mathbf{r} = (0, 0, 1)^T$ by performing the unitary time evolution $U_\alpha(\mathbf{n})$ we can transform this state in any other state on the Bloch sphere with Bloch vector $\mathbf{r}'$ by just choosing some specific $\alpha, \mathbf{n}$ such that the corresponding rotation of the Bloch vector $R_\alpha(\mathbf{n})(0, 0, 1)^T = \mathbf{r}'$. This will be extremely useful later on when discussing applications of entanglement.

5
Bipartite systems

LECTURE 9

Many interesting features of QI arise when considering bipartite systems. This means that we consider a system with Hilbert space $\mathcal{H}$ which can be separated (partitioned) into two subsystems, A and B, described by Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. Conventionally we introduce two characters, Alice who has system A and Bob who has system B. We assume that they each have full control over their own systems, but no direct control over the other's system. In particular we assume Alice can perform any time-evolution in system A and can make any measurement in that system. This means that she can choose any Hamiltonian for her system, e.g. by rotating the system, choosing to apply an electric or magnetic field etc. Likewise, Bob has full control of system B. We say then that each can perform arbitrary *Local Operations* (LO).

We may also allow Alice and Bob to communicate through classical channels, i.e. *Classical Communications* (CC). This means that they can send classical bits to each other. This includes speaking on the phone or sending texts and emails, but we often want to quantify the amount of information exchanged so we count the number of bits transferred. (This counting does not include any pre-agreed protocol which is required to interpret the data transferred. I.e. if Alice sends a single bit to Bob with value 1, we assume he knows whether this means "I measured X and got value 1 rather than 0" or "I decided to perform unitary transformation U rather than V".) Together if Alice and Bob can both perform LO and communicate classically, we say they can perform LOCC operations.

An additional possibility is that Alice and Bob can communicate through quantum channels. This means that they can send qubits to each other. This is more powerful than classical communication. In fact with unlimited capacity for quantum communication, they can perform arbitrary operations on the whole system. A simple argument is that Alice could just send her whole quantum system to Bob. He could then do anything on the full system and then send subsystem A back to Alice. However, we can focus on specific issues such as what can be done by sending a single qubit compared to a single bit.

5.1 Tensor Products

When we say that we consider a system with Hilbert space $\mathcal{H}$ which can be separated (partitioned) into two (or more) subsystems in mathematical language we are saying that our Hilbert space can be written as a *tensor product* of two (or more) Hilbert spaces and we will write it as $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$.

The tensor product of two vector spaces is a new vector space and this is a construction that could be discussed in a purely linear algebra setup without ever mentioning quantum mechanics similar to what you have seen when discussing the direct sum $\mathcal{H}_A \oplus \mathcal{H}_B$.

In this module we will not be needing the most general definition of tensor product so we will just start from the vectors of $\mathcal{H}_A$ and $\mathcal{H}_B$ to construct the vectors of this new vector space $\mathcal{H}_A \otimes \mathcal{H}_B$ and understand how the usual linearity properties work in this extended space.

With this in mind let us denote $|\psi_1\rangle \in \mathcal{H}_A$ and $|\psi_2\rangle \in \mathcal{H}_B$ then the vector $|\psi_1\rangle \otimes |\psi_2\rangle \in \mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. We can write a basis $\{|i\rangle \otimes |m\rangle\}$ for $\mathcal{H}$ in terms of a basis $\{|i\rangle\}$ for $\mathcal{H}_A$ and a basis $\{|m\rangle\}$ for $\mathcal{H}_B$. Just by counting how many basis elements we have if $\mathcal{H}_A$ has dimension N_A and $\mathcal{H}_B$ has dimension N_B we have that $\mathcal{H}$ has dimension $N_A N_B$.

The tensor product has various linear properties

- $c\Big(|\psi\rangle \otimes |\phi\rangle\Big) = \Big(c\,|\psi\rangle\Big) \otimes |\phi\rangle = |\psi\rangle \otimes \Big(c\,|\phi\rangle\Big)$
- $a\,|\psi_1\rangle \otimes |\phi\rangle + b\,|\psi_2\rangle \otimes |\phi\rangle = \Big(a\,|\psi_1\rangle + b\,|\psi_2\rangle\Big) \otimes |\phi\rangle$
- $a\,|\psi\rangle \otimes |\phi_1\rangle + b\,|\psi\rangle \otimes |\phi_2\rangle = |\psi\rangle \otimes \Big(a\,|\phi_1\rangle + b\,|\phi_2\rangle\Big)$

Note in particular that in order to simplify the sum of two tensors we must have that either the first vector is the same or the second one is, if we find an expression of the form $|\psi_1\rangle \otimes |\phi_1\rangle + |\psi_2\rangle \otimes |\phi_2\rangle$ we have to leave it like that.

The inner products of $\mathcal{H}_A$ and $\mathcal{H}_B$ induce an inner product on the tensor product space so that the inner product of $|\psi_1\rangle \otimes |\phi_1\rangle$ with $|\psi_2\rangle \otimes |\phi_2\rangle$ in $\mathcal{H}$ is defined by the inner products in $\mathcal{H}_A$ and $\mathcal{H}_B$ as:

$$\Big(\langle\psi_1| \otimes \langle\phi_1|\Big)\Big(|\psi_2\rangle \otimes |\phi_2\rangle\Big) = \langle\psi_1|\psi_2\rangle\ \langle\phi_1|\phi_2\rangle\ .$$

If a pure state can be written in the form $|\psi\rangle \otimes |\phi\rangle$ we say it is a *separable* state. However, a typical state is a linear combination of such states and is not separable. We say a non-separable state is *entangled*.

States of a bipartite system which can be written as $|\psi\rangle \otimes |\phi\rangle$ are *separable*, all others are *entangled*.

For example suppose that the basis $\{|i\rangle\}$ for $\mathcal{H}_A$ is orthonormal as well as the basis $\{|m\rangle\}$ for $\mathcal{H}_B$, then the basis $\{|i\rangle \otimes |m\rangle\}$ for $\mathcal{H}$ will also be orthonormal. We can then write a separable vector

$$|\psi\rangle \otimes |\phi\rangle = \left(\sum_i a_i\,|i\rangle\right) \otimes \left(\sum_m b_m\,|m\rangle\right) = \sum_{i,m} a_i b_m\,|i\rangle \otimes |m\rangle\ .$$

However the most general vector $|\Psi\rangle \in \mathcal{H}$ will be written as

The components of a bipartite state $|\Psi\rangle$ on a basis $\{|i\rangle \otimes |m\rangle\}$ are two-index coefficients c_{im} (tensor components).

$$|\Psi\rangle = \sum_{i,m} c_{im} |i\rangle \otimes |m\rangle \ ,$$

for some complex numbers $c_{im} \in \mathbb{C}$ which generically cannot be written as $a_i b_m$. We will see later on a quantity that will tell us immediately if a state is entangled or separable.

If we have the orthonormal basis $\{|i\rangle \otimes |m\rangle\}$ the inner product between two general states can then be computed as

$$\begin{aligned}\langle \Phi | \Psi \rangle &= \left(\sum_{j,n} d^*_{jn} \langle j| \otimes \langle n| \right) \left(\sum_{i,m} c_{im} |i\rangle \otimes |m\rangle \right) \\ &= \sum_{i,j,m,n} d^*_{jn} c_{im} \langle j | i \rangle \langle n | m \rangle = \sum_{i,m} d^*_{im} c_{im} \ ,\end{aligned}$$

where we used the orthonormality of the $|i\rangle$ and $|m\rangle$.

The inner product between two states of a bipartite system, expressed in components, equals $\sum_{i,m} d^*_{im} c_{im}$.

We can construct mixed ensemble of separable and entangled states however we need first to understand how linear operators work on tensor product spaces (see next section).

Once we have the tensor product of two vector spaces $\mathcal{H}_A \otimes \mathcal{H}_B$ we can consider more and more tensors, i.e. $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. For example the Hilbert space of a N-qubits system is the 2^N dimensional Hilbert space

$$H_{\text{N-qubits}} = \mathcal{H}_q^{\otimes N} \ ,$$

where we simply mean the tensor product $\mathcal{H}_q \otimes \mathcal{H}_q \otimes ... \otimes \mathcal{H}_q$ of N copies of a single qubit system.

Quantum computing will be the analysis of algorithms performed on the space of N-qubits seen as quantum circuits built out of gates (unitary operators) acting on a single qubit state or on pairs of qubits, analogues of the classical logical gates NOT, AND, OR, XOR acting on usual strings of bits, e.g. 01101000 01101001,... .

> **Example:**
>
> Consider the space of 3-qubits
>
> $$\mathcal{H} = \mathcal{H}_q \otimes \mathcal{H}_q \otimes \mathcal{H}_q = \text{span}\{|000\rangle, |001\rangle, |010\rangle, |011\rangle, |100\rangle, |101\rangle, |110\rangle, |111\rangle\}$$
>
> where we used the shorthand notation $|000\rangle = |0\rangle \otimes |0\rangle \otimes |0\rangle$ and so on.
>
> The operator $\hat{I} \otimes \sigma_1 \otimes \hat{I}$ will act on the second qubit and leave the first two qubits invariant, furthermore we know that $\sigma_1 |0\rangle = |1\rangle$ and $\sigma_1 |1\rangle = |0\rangle$ hence σ_1 acts on this basis as the usual NOT logical gate where NOT $0 = \bar{0} = 1$ and NOT $1 = \bar{1} = 0$. We can then write the action of this operator on the general 3-qubit state $|xyz\rangle$ with $x, y, z \in \{0, 1\}$ as
>
> $$\hat{I} \otimes \sigma_1 \otimes \hat{I} |xyz\rangle = |x\bar{y}z\rangle \ .$$
>
> Similarly an operator of the form $\sum_J \hat{A}_J \otimes \hat{I} \otimes \hat{B}_J$ will act both on the first and third qubit while leaving the second qubit unchanged.

5.2 Linear Operators and Local Unitary Operations

LECTURE 11

Linear operators on $\mathcal{H}$ can be written as linear combinations of operators of the form $\hat{A}\otimes\hat{B}$. (As for separable states, a general linear operator cannot be written in this way as a single tensor product.) Such operators act on separable states as

$$\hat{A}\otimes\hat{B}\,|\psi\rangle\otimes|\phi\rangle=\left(\hat{A}\,|\psi\rangle\right)\otimes\left(\hat{B}\,|\phi\rangle\right).$$

By linearity this defines the action of a general linear operator on an arbitrary state in $\mathcal{H}$.

As for addition of tensor product of vectors, also for operators we cannot simplify $\hat{A}\otimes\hat{B}+\hat{C}\otimes\hat{D}$ in general, however if either the first factors are the same, or the second factors are, we have

$$\hat{A}\otimes\hat{B}+\hat{C}\otimes\hat{B}=(\hat{A}+\hat{C})\otimes\hat{B}\,,\qquad \hat{A}\otimes\hat{B}+\hat{A}\otimes\hat{D}=\hat{A}\otimes(\hat{B}+\hat{D})\,.$$

Note that standard operation or properties of the operators $\hat{A}$ and $\hat{B}$ descend to their tensor product, for example

$$\begin{aligned}(\hat{A}\otimes\hat{B})^{\dagger}&=\hat{A}^{\dagger}\otimes\hat{B}^{\dagger}\,,\\(\hat{A}\otimes\hat{B})(\hat{C}\otimes\hat{D})&=(\hat{A}\hat{C}\otimes\hat{B}\hat{D})\,,\\\mathrm{Tr}_{\mathcal{H}_A\otimes\mathcal{H}_B}(\hat{A}\otimes\hat{B})&=\mathrm{Tr}_{\mathcal{H}_A}(\hat{A})\,\mathrm{Tr}_{\mathcal{H}_B}(\hat{B})\,,\end{aligned}$$

or

- the tensor product of two unitary operators is unitary;
- the tensor product of two hermitian operators is hermitian;
- the tensor product of two positive operators is positive;
- the tensor product ot two projectors is a projector.

If we have a bipartite system and consider only local unitary operations then Alice and Bob can each perform only very restricted unitary transformations of the form $\hat{U}_A\otimes\hat{I}$ for Alice and $\hat{I}\otimes\hat{U}_B$ for Bob. Note that these two operators commute

$$\begin{aligned}[\hat{U}_A\otimes\hat{I},\hat{I}\otimes\hat{U}_B]&=\left(\hat{U}_A\otimes\hat{I}\right)\left(\hat{I}\otimes\hat{U}_B\right)-\left(\hat{I}\otimes\hat{U}_B\right)\left(\hat{U}_A\otimes\hat{I}\right)\\&=\left(\hat{U}_A\hat{I}\right)\otimes\left(\hat{I}\,\hat{U}_B\right)-\left(\hat{I}\,\hat{U}_A\right)\otimes\left(\hat{U}_B\hat{I}\right)\\&=\hat{U}_A\otimes\hat{U}_B-\hat{U}_A\otimes\hat{U}_B=0\,,\end{aligned}$$

so Alice and Bob independently transform their own systems, and their product is $\hat{U}_A\otimes\hat{U}_B$ which can be seen easily from

$$\begin{aligned}(\hat{I}\otimes\hat{U}_B)(\hat{U}_A\otimes\hat{I})\,|\psi\rangle\otimes|\phi\rangle&=(\hat{I}\otimes\hat{U}_B)(\hat{U}_A\,|\psi\rangle)\otimes(\hat{I}\,|\phi\rangle)=(\hat{I}\hat{U}_A\,|\psi\rangle)\otimes(\hat{U}_B\hat{I}\,|\phi\rangle)\\&=(\hat{U}_A\,|\psi\rangle)\otimes(\hat{U}_B\,|\phi\rangle)=(\hat{U}_A\otimes\hat{U}_B)\,|\psi\rangle\otimes|\phi\rangle\ ,\end{aligned}$$

so we see that together Alice and Bob can only transform the system by unitary operators of this restricted form.

Local operations are of the form $\hat{U}_A\otimes\hat{U}_B$, while general unitary operators on the bipartite Hilbert space cannot be simplified to this form.

Local operations cannot turn a separable state into an entangled state.

One very important result is that if Alice and Bob start with a separable state $|\psi\rangle\otimes|\phi\rangle$ then the most general unitary transformation they can perform using LO will produce another separable state, $\hat{U}_A\,|\psi\rangle\otimes\hat{U}_B\,|\phi\rangle$. I.e. they cannot create an entangled state from a separable state.

> **Example:**
>
> Let us compute the exponential of the operators $\hat{A}\otimes\hat{I}$ and $\hat{I}\otimes\hat{B}$. We have
>
> $$e^{\hat{A}\otimes\hat{I}}=\sum_{n=0}^{\infty}\frac{(\hat{A}\otimes\hat{I})^n}{n!}=\sum_{n=0}^{\infty}\frac{\hat{A}^n\otimes\hat{I}^n}{n!}=e^{\hat{A}}\otimes\hat{I}\,,$$
> $$e^{\hat{I}\otimes\hat{B}}=\sum_{n=0}^{\infty}\frac{(\hat{I}\otimes\hat{B})^n}{n!}=\sum_{n=0}^{\infty}\frac{\hat{I}^n\otimes\hat{B}^n}{n!}=\hat{I}\otimes e^{\hat{B}}\,.$$
>
> Hence we have
>
> $$e^{\hat{A}\otimes\hat{I}}e^{\hat{I}\otimes\hat{B}}=\left(e^{\hat{A}}\otimes\hat{I}\right)\left(\hat{I}\otimes e^{\hat{B}}\right)=e^{\hat{A}}\otimes e^{\hat{B}}\,,$$
>
> in particular notice that for general operators $\hat{A}\,,\hat{B}$ the exponential of $\hat{A}\otimes\hat{B}$, i.e. $e^{\hat{A}\otimes\hat{B}}$ is different from $e^{\hat{A}}\otimes e^{\hat{B}}$ since
>
> $$e^{\hat{A}\otimes\hat{B}}=\sum_{n=0}^{\infty}\frac{(\hat{A}\otimes\hat{B})^n}{n!}=\sum_{n=0}^{\infty}\frac{\hat{A}^n\otimes\hat{B}^n}{n!}\,,$$
> $$e^{\hat{A}}\otimes e^{\hat{B}}=\left(\sum_{n_1=0}^{\infty}\frac{\hat{A}^{n_1}}{n_1!}\right)\left(\sum_{n_2=0}^{\infty}\frac{\hat{B}^{n_2}}{n_2!}\right)=\sum_{n_1,n_2=0}^{\infty}\frac{\hat{A}^{n_1}\otimes\hat{B}^{n_2}}{n_1!\,n_2!}\,.$$

> **Example:**
>
> Compute the commutator of $\hat{A}_1\otimes\hat{B}$ with $\hat{A}_2\otimes\hat{I}$. We have
>
> $$\begin{aligned}\left(\hat{A}_1\otimes\hat{B}\right)\left(\hat{A}_2\otimes\hat{I}\right)&=\left(\hat{A}_1\hat{A}_2\right)\otimes\hat{B}\,,\\ \left(\hat{A}_2\otimes\hat{I}\right)\left(\hat{A}_1\otimes\hat{B}\right)&=\left(\hat{A}_2\hat{A}_1\right)\otimes\hat{B}\,,\\ \left[\hat{A}_1\otimes\hat{B},\hat{A}_2\otimes\hat{I}\right]&=\left(\hat{A}_1\otimes\hat{B}\right)\left(\hat{A}_2\otimes\hat{I}\right)-\left(\hat{A}_2\otimes\hat{I}\right)\left(\hat{A}_1\otimes\hat{B}\right)\\ &=\left(\hat{A}_1\hat{A}_2\right)\otimes\hat{B}-\left(\hat{A}_2\hat{A}_1\right)\otimes\hat{B}\\ &=\left[\hat{A}_1,\hat{A}_2\right]\otimes\hat{B}\,.\end{aligned}$$
>
> Similarly we can easily obtain
>
> $$\left[\hat{A}\otimes\hat{B}_1,\hat{I}\otimes\hat{B}_2\right]=\hat{A}\otimes\left[\hat{B}_1,\hat{B}_2\right]\,.$$
>
> Note in particular that $\left[\hat{A}_1\otimes\hat{B}_1,\hat{A}_2\otimes\hat{B}_2\right]$ is in general different from $\left[\hat{A}_1,\hat{A}_2\right]\otimes\left[\hat{B}_1,\hat{B}_2\right]$ as we can see by expanding everything out
>
> $$\begin{aligned}\left[\hat{A}_1\otimes\hat{B}_1,\hat{A}_2\otimes\hat{B}_2\right]&=\left(\hat{A}_1\otimes\hat{B}_1\right)\left(\hat{A}_2\otimes\hat{B}_2\right)-\left(\hat{A}_2\otimes\hat{B}_2\right)\left(\hat{A}_1\otimes\hat{B}_1\right)\\ &=\hat{A}_1\hat{A}_2\otimes\hat{B}_1\hat{B}_2-\hat{A}_2\hat{A}_1\otimes\hat{B}_2\hat{B}_1\,,\\ \left[\hat{A}_1,\hat{A}_2\right]\otimes\left[\hat{B}_1,\hat{B}_2\right]&=\left(\hat{A}_1\hat{A}_2-\hat{A}_2\hat{A}_1\right)\otimes\left(\hat{B}_1\hat{B}_2-\hat{B}_2\hat{B}_1\right)\\ &=\hat{A}_1\hat{A}_2\otimes\hat{B}_1\hat{B}_2-\hat{A}_1\hat{A}_2\otimes\hat{B}_2\hat{B}_1-\hat{A}_2\hat{A}_1\otimes\hat{B}_1\hat{B}_2+\hat{A}_2\hat{A}_1\otimes\hat{B}_2\hat{B}_1\,.\end{aligned}$$

5.2.1 Density matrix for tensor products

We can now consider the density matrix associated to a separable state $|\Psi\rangle = |\psi\rangle \otimes |\phi\rangle$ which is simply

$$\hat{\rho} = |\Psi\rangle\langle\Psi| = \Big(|\psi\rangle\langle\psi|\Big) \otimes \Big(|\phi\rangle\langle\phi|\Big) = \hat{\rho}_A \otimes \hat{\rho}_B\,,$$

where $\hat{\rho}_A = |\psi\rangle\langle\psi|$ is just the density matrix for system A and $\hat{\rho}_B = |\phi\rangle\langle\phi|$ the density matrix for system B.

A mixed ensemble of separable states is then given by

$$\hat{\rho} = \sum_n p_n\, \hat{\rho}_A^{(n)} \otimes \hat{\rho}_B^{(n)}\,,$$

where $\{\hat{\rho}_A^{(n)}\}$ are mixed or pure states of system A, while $\{\hat{\rho}_B^{(n)}\}$ are mixed or pure states of system B and as always $p_n \geq 0$ such that $\sum_n p_n = 1$.

We will say that a mixed state is separable if and only if it is an ensemble of separable states, entangled otherwise. We will not be focusing particularly on mixed states in bipartite systems because they can be seen as arising from a pure state in a larger system via the process of reduced density matrix and partial trace (see later).

5.3 Matrix Representation

Since $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is once again a vector space with a complex inner product, we have that states and operators in $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ can be expressed as vectors and matrices. To express the tensor product of two column vectors or matrices we use the convention that the first term gives the block structure while the second specifies the detail of the individual blocks up to multiplication by the appropriate constant from the first. This is simplest to explain with examples

$$\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} \otimes \begin{pmatrix} 3 \\ 4 \end{pmatrix} \to \begin{pmatrix} 1\begin{pmatrix} 3 \\ 4 \end{pmatrix} \\ 2\begin{pmatrix} 3 \\ 4 \end{pmatrix} \\ 5\begin{pmatrix} 3 \\ 4 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 6 \\ 8 \\ 15 \\ 20 \end{pmatrix}$$

where the middle expression is just given to indicate the structure.

This process can be understood by ordering the basis elements of $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ starting from the basis of $\mathcal{H}_A$ and $\mathcal{H}_B$ with the following order

$$\begin{aligned}\mathcal{H} = \text{span}\Bigg\{ &\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ &\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Bigg\}\,,\end{aligned}$$

and now we represent this basis with the standard basis in the order given, i.e.

$$\begin{pmatrix}1\\0\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}\rightarrow\begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix}\,,\ \dots\ ,\begin{pmatrix}0\\0\\1\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}\rightarrow\begin{pmatrix}0\\0\\0\\0\\0\\1\end{pmatrix}\,,$$

which produces exactly the procedure outlined above with an example.

Similarly for matrices

$$\begin{pmatrix}1&0&0\\0&0&2\\0&-3&0\end{pmatrix}\otimes\begin{pmatrix}0&1\\-1&2\end{pmatrix}\rightarrow\begin{pmatrix}0&1&0&0&0&0\\-1&2&0&0&0&0\\0&0&0&0&0&2\\0&0&0&0&-2&4\\0&0&0&-3&0&0\\0&0&3&-6&0&0\end{pmatrix}\,.$$

In the above examples the first Hilbert space has dimension 3 while the second has dimension 2 but obviously similar relations hold for arbitrary (finite) dimensional Hilbert spaces.

Assuming we use orthonormal basis for the original Hilbert spaces, these vector and matrix representations correspond to an orthonormal basis for the tensor product Hilbert space.

Note that both the most general vector and the most general linear operator are not of the simple form $|\psi\rangle\otimes|\phi\rangle$ and $\hat{A}\otimes\hat{B}$ but rather $\sum_{a,b}|\psi_a\rangle\otimes|\phi_b\rangle$ and $\sum_{I,J}\hat{A}_I\otimes\hat{B}_J$.

> **Example:**
>
> Let $\mathcal{H}=\mathcal{H}_q\otimes\mathcal{H}_q$ be the Hilbert space of two qubits, i.e. a four dimensional Hilbert space. If possible write the linear operator on $\mathcal{H}$
>
> $$O=\begin{pmatrix}1&0&0&2\\0&1&2&0\\0&2&1&0\\2&0&0&1\end{pmatrix}$$
>
> as $A\otimes B$ where A,B are linear operators on $\mathcal{H}_q$, i.e. 2×2 matrices. First we notice that
>
> $$O=\mathbb{1}_4+\begin{pmatrix}0&0&0&2\\0&0&2&0\\0&2&0&0\\2&0&0&0\end{pmatrix}=\mathbb{1}_2\otimes\mathbb{1}_2+\begin{pmatrix}0&2\\2&0\end{pmatrix}\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}=\mathbb{1}_2\otimes\mathbb{1}_2+2\,\sigma_1\otimes\sigma_1\,,$$
>
> which cannot be simplified any further so $O\neq A\otimes B$.

> **Example:**
>
> Let us consider again $\mathcal{H}=\mathcal{H}_q\otimes\mathcal{H}_q$ and define
>
> $$U=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix}\,.$$

First check that U is a unitary operator. Let us write it as $A_1 \otimes B_1 + A_2 \otimes B_2$, to do this we write

$$\begin{aligned}
U &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
&= \frac{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}}{2} \otimes \mathbb{1}_2 + \frac{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}}{2} \otimes \sigma_1 = \left(\frac{\mathbb{1}_2 + \sigma_3}{2} \right) \otimes \mathbb{1}_2 + \left(\frac{\mathbb{1}_2 - \sigma_3}{2} \right) \otimes \sigma_1 \,.
\end{aligned}$$

We can now understand why this operator implements the quantum logical gate called *Controlled NOT* (CNOT) where the first qubit controls what we do to the second qubit, if the first qubit is $|0\rangle$ we leave the second qubit invariant, otherwise if the first qubit is $|1\rangle$ we take the NOT of the second qubit $|\bar{y}\rangle = \sigma_1 |y\rangle$.

As always to understand what an operator does we just need to check what U does on the four basis states $\mathcal{H} = \text{span}\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$.

First of all check that

$$\left(\frac{\mathbb{1}_2 + \sigma_3}{2} \right) |0\rangle = |0\rangle \,, \qquad \left(\frac{\mathbb{1}_2 + \sigma_3}{2} \right) |1\rangle = 0 \,,$$

and similarly

$$\left(\frac{\mathbb{1}_2 - \sigma_3}{2} \right) |0\rangle = 0 \,, \qquad \left(\frac{\mathbb{1}_2 - \sigma_3}{2} \right) |1\rangle = |1\rangle \,,$$

so the first operators appearing in U acting on the first qubit behave as a *control*:

-If the first qubit is $|0\rangle$ we only need to consider $\left(\frac{\mathbb{1}_2+\sigma_3}{2}\right) \otimes \mathbb{1}_2$, which acts as the identity on the second qubit;

-while if the first qubit is $|1\rangle$ we only need to consider $\left(\frac{\mathbb{1}_2-\sigma_3}{2}\right) \otimes \sigma_1$ which acts as the logical gate NOT on the second qubit and we have

$$\begin{aligned}
U|00\rangle &= \left(\frac{\mathbb{1}_2 + \sigma_3}{2} \right) |0\rangle \otimes \mathbb{1}_2 |0\rangle + && = |00\rangle \,, \\
U|01\rangle &= \left(\frac{\mathbb{1}_2 + \sigma_3}{2} \right) |0\rangle \otimes \mathbb{1}_2 |1\rangle + && = |01\rangle \,, \\
U|10\rangle &= && + \left(\frac{\mathbb{1}_2 - \sigma_3}{2} \right) |1\rangle \otimes \sigma_1 |0\rangle = |11\rangle \,, \\
U|11\rangle &= && + \left(\frac{\mathbb{1}_2 - \sigma_3}{2} \right) |1\rangle \otimes \sigma_1 |1\rangle = |10\rangle \,.
\end{aligned}$$

Check that this is exactly the same action as the 4×4 matrix written above has on the four standard basis vectors $\{(1,0,0,0)^T, ..., (0,0,0,1)^T\}$.

5.4 Local Measurements

LECTURE 12

If Alice and Bob perform measurements on their own systems, they can do so using self-adjoint operators of the form $\hat{F} = \hat{F}_A \otimes \hat{I}$ for Alice and $\hat{G} = \hat{I} \otimes \hat{G}_B$ for Bob. Assuming for simplicity no degeneracy for the spectrum of $\hat{F}_A$ or $\hat{G}_B$ within each subsystem, these operators have projection operators $\hat{F}_{Ai} = |i\rangle\langle i|$ and $\hat{G}_{Bm} = |m\rangle\langle m|$ within Alice's and Bob's subsystems respectively.

In the full system the operators are degenerate, with degeneracy given by the dimension of the other subsystem, i.e. by the dimension of $\mathcal{H}_B$ for Alice's observable, and by that of $\mathcal{H}_A$ for Bob's. The corresponding projection operators in the full system are $\hat{F}_i = \hat{F}_{Ai} \otimes \hat{I}$ and $\hat{G}_m = \hat{I} \otimes \hat{G}_{Bm}$.

Using the spectral decomposition of the identity we can write these two projectors as

$$\hat{F}_i = \sum_n |i\rangle \langle i| \otimes |n\rangle \langle n| \; , \qquad \hat{G}_m = \sum_j |j\rangle \langle j| \otimes |m\rangle \langle m| \; .$$

It is clear then that the eigenspace for the operator $\hat{F}_A$ which was spanned by $|i\rangle$ in $\mathcal{H}_A$ is now becoming degenerate for the operator $\hat{F} = \hat{F}_A \otimes \hat{I}$ in $\mathcal{H}_A \otimes \mathcal{H}_B$ spanned by all the vectors of the form $|i\rangle \otimes |\phi\rangle$ with $|\phi\rangle \in \mathcal{H}_B$.

Since $\left[\hat{F}, \hat{G}\right] = 0$ the measurements are compatible so Alice and Bob can both measure and the final state will be in a simultaneous eigenstate of $\hat{F}$ and $\hat{G}$. The outcomes of the measurements will be some pair (f_i, g_m) and the probability of this outcome is independent of whether Alice or Bob measures first – in fact they could also measure simultaneously (or with spacelike separation so that no signal could travel between them to allow one measurement to potentially affect the other.) In this sense, as for local unitary transformations, they can consider their systems to be isolated from each other.

Suppose in fact that the state is in a pure state and let us assume for now separable as well

$$|\Psi\rangle = |\psi\rangle \otimes |\phi\rangle = \sum_{i\,,m} \alpha_i \beta_m \, |i\rangle \otimes |m\rangle \; ,$$

where the coefficients $\alpha_i, \beta_m \in \mathbb{C}$ and $\{|i\rangle\}$, $\{|m\rangle\}$ form an orthonormal basis for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively.

As always the states $|\psi\rangle$ and $|\phi\rangle$ are normalised which means $\sum_i |\alpha_i|^2 = \sum_m |\beta_m|^2 = 1$, this implies that if we define the combination $\alpha_i \beta_m = \gamma_{im}$ then the condition $\langle \Psi \,|\Psi\rangle = 1$ imposes $\sum_{i\,,m} |\gamma_{im}|^2 = 1$.

If Alice measures $\hat{F}$ she will obtain outcome f_j with probability $|\alpha_j|^2 = \sum_m |\gamma_{jm}|^2$ and the system will then collapse to the state

$$\sum_m \beta_m \, |j\rangle \otimes |m\rangle = |j\rangle \otimes |\phi\rangle \; .$$

If Bob then measures $\hat{G}$ and obtains outcome g_n with probability $|\beta_n|^2 = \sum_i |\gamma_{im}|^2$ we have that the final state becomes $|j\rangle \otimes |m\rangle$. The probability would be exactly the same if Bob had measured first, except the intermediate state would have been $|\psi\rangle \otimes |n\rangle$. Overall the combined measurements of $\hat{F}$ and $\hat{G}$ with outcomes (f_j, g_n) have probability $|\gamma_{jn}|^2$ which is the product of the two probabilities.

We can also rewrite everything in operator formalism. We just need to remember that the operators $\hat{F}_{Ai} = |i\rangle \langle i|$ are mutually orthogonal projectors onto the eigenspace span$\{|i\rangle\} \subseteq \mathcal{H}_A$. The probability for Alice to measure f_i is

$$|\,\langle i\,|\psi\rangle\,|^2 = \mathrm{Tr}\left(\hat{\rho}_A \, \hat{F}_{Ai}\right) \; ,$$

and after measuring $\hat{F}_A$ and finding f_i the state of Alice has collapsed to the normalised state

$$|\psi\rangle \to |i\rangle = \frac{1}{|\langle i\,|\psi\rangle\,|}\hat{F}_{Ai}\,|\psi\rangle = \frac{1}{\sqrt{\mathrm{Tr}\left(\hat{\rho}_A\,\hat{F}_{Ai}\right)}}\hat{F}_{Ai}\,|\psi\rangle \ ,$$

or equivalently

$$\hat{\rho}_A \to \frac{1}{\mathrm{Tr}\left(\hat{\rho}_A\,\hat{F}_{Ai}\right)}\hat{F}_{Ai}\,\hat{\rho}_A\hat{F}_{Ai}\,,$$

where we have intensively used the fact that $\hat{F}_{Ai}$ is a projector, i.e. $\hat{F}^\dagger_{Ai} = \hat{F}_{Ai}$ and $\hat{F}^2_{Ai} = \hat{F}_{Ai}$.

For a bipartite system we just need to repeat this analysis while carrying along the road Bob's system. Clearly when Alice measures the observable $\hat{F}_A$ on her system she does not perform any operation on Bob's (LO) so we define

$$\hat{F}_i = \hat{F}_{Ai} \otimes \hat{I}\,.$$

If we have prepared the separable state $|\Psi\rangle = |\psi\rangle \otimes |\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ then we can use the density matrix

$$\hat{\rho} = |\Psi\rangle\,\langle\Psi| = |\psi\rangle\,\langle\psi| \otimes |\phi\rangle\,\langle\phi| = \hat{\rho}_A \otimes \hat{\rho}_B\,.$$

Now if Alice measures $\hat{F} = \hat{F}_A \otimes \hat{I}$ she will obtain outcome f_i with probability

$$\mathrm{Tr}_{A\otimes B}\left(\hat{\rho}\,\hat{F}_i\right) = \mathrm{Tr}_A\left(\hat{\rho}_A\,\hat{F}_{Ai}\right) \ ,$$

and after that her wave function will have collapsed while Bob' state will be unchanged

$$\hat{\rho} \to \frac{1}{\mathrm{Tr}_A\left(\hat{\rho}_A\,\hat{F}_{Ai}\right)}\hat{F}_{Ai}\,\hat{\rho}_A\hat{F}_{Ai} \otimes \hat{\rho}_B = \frac{1}{\mathrm{Tr}_{A\otimes B}\left(\hat{\rho}\,\hat{F}_i\right)}\hat{F}_i\,\hat{\rho}\,\hat{F}_i\,,$$

where we used the projector $\hat{F}_i$ defined on the whole bipartite system.

Note in particular that

$$\hat{F}_i = \hat{F}_{Ai} \otimes \hat{I} = \sum_m \left(|i\rangle \otimes |m\rangle\right) \otimes \left(\langle i| \otimes \langle m|\right) \ ,$$

where we have used the spectral decomposition of $\hat{I}$ for Bob system to make manifest the fact that although the eigenspace corresponding to the eigenvalue f_i was non-degenerate in $\mathcal{H}_A$, as soon as we consider a bipartite system it immediately becomes degenerate since any vector of the form $|i\rangle \otimes |\phi\rangle$ with $|\phi\rangle \in \mathcal{H}_B$ is an eigenvector of $\hat{F} \otimes \hat{I}$ with same eigenvalue f_i (see comment at the beginning of this section).

Now similarly if Bob measures $\hat{G}_B$ with spectrum $\{g_n, |n\rangle\}$ he will obtain outcome g_m with probability

$$\mathrm{Tr}_B\left(\hat{\rho}_B\,\hat{G}_{Bm}\right) = \mathrm{Tr}_{A\otimes B}\left(\hat{\rho}\,\hat{G}_m\right) \ ,$$

where we defined $\hat{G}_{Bm} = |m\rangle \langle m|$ and $\hat{G}_m = \hat{I} \otimes \hat{G}_{Bm}$. After measurement the wave function will collapse to

$$\hat{\rho} \to \frac{1}{\mathrm{Tr}_{A\otimes B}\left(\hat{\rho}\,\hat{G}_m\right)} \hat{G}_m\,\hat{\rho}\,\hat{G}_m\,.$$

As above it does not matter who measures first (only the intermediate state will change), if we measure $\hat{F} = \hat{F}_A \otimes \hat{I}$ and $\hat{G} = \hat{I} \otimes \hat{G}_B$ the outcome (f_i, g_m) will be measured with probability

$$\mathrm{Tr}_{A\otimes B}\left(\hat{\rho}\,\hat{P}_{im}\right)\;,$$

with $\hat{P}_{im} = \hat{F}_{Ai} \otimes \hat{G}_{Bm} = |i\rangle \langle i| \otimes |m\rangle \langle m|$.

The state will then collapse to

$$\hat{\rho} \to \frac{1}{\mathrm{Tr}_{A\otimes B}\left(\hat{\rho}\,\hat{P}_{im}\right)} \hat{P}_{im}\,\hat{\rho}\,\hat{P}_{im} = |i\rangle \otimes |m\rangle\;.$$

Let us now repeat the same analysis but for an entangled state. Suppose the bipartite system $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is prepared in the state

$$|\Psi\rangle = \sum_{i,m} \gamma_{im}\,|i\rangle \otimes |m\rangle\;,$$

where the normalised coefficients $\gamma_{im} \in \mathbb{C}$ are not of the form $\gamma_{im} \neq \alpha_i \beta_m$ as above, i.e. the state is an entangled one.

Let us still define the coefficients

$$\alpha_j = \sqrt{\sum_m |\gamma_{jm}|^2}\,, \qquad \beta_n = \sqrt{\sum_i |\gamma_{in}|^2}\,,$$

allowing us to define the two auxiliary normalised states

$$\begin{aligned} |\psi_n\rangle &= \frac{1}{\beta_n} \sum_i \gamma_{in}\,|i\rangle \in \mathcal{H}_A\,, \\ |\phi_j\rangle &= \frac{1}{\alpha_j} \sum_m \gamma_{jm}\,|m\rangle \in \mathcal{H}_B\,, \end{aligned}$$

excluding values of n and j for which $\beta_n = 0$ or $\alpha_j = 0$. (Little exercise for you check that these states are indeed normalised.)

We can then rewrite the state $|\Psi\rangle$ as

$$\begin{aligned} |\Psi\rangle &= \sum_{i,m} \gamma_{im}\,|i\rangle \otimes |m\rangle\;, \\ &= \sum_i \alpha_i\,|i\rangle \otimes |\phi_i\rangle\;, \\ &= \sum_n \beta_m\,|\psi_m\rangle \otimes |m\rangle\;. \end{aligned}$$

If Alice measures $\hat{F}$ she will have outcome f_i with probability $|\alpha_i|^2$ and after the measurement the state will have collapsed to

$$|\Psi\rangle \to \hat{F}_i\,|\Psi\rangle = (\hat{F}_{Ai} \otimes \hat{I})\,|\Psi\rangle \sim |i\rangle \otimes |\phi_i\rangle \ .$$

The key difference from the previous, separable case is that starting from an entangled state after Alice measurement we obtain a separable one with $|i\rangle \in \mathcal{H}_A$ and $|\phi_i\rangle \in \mathcal{H}_B$ but Bob' state depends on the result of Alice measurement!

This is the novelty of quantum mechanics: we say that quantum mechanics is *non-local*! Local measurements, for example Alice measuring a spin in her laboratory in New York, can have non-local effects, i.e. changing Bob' state who lives on Mars. We will shortly see that this "spooky action at a distance", as Einstein used to call it, will not allow us to communicate faster than the speed of light, i.e. it will not violate causality. The key point is that if both Alice and Bob know the full initial state $|\Psi\rangle$, then after measuring $\hat{F}$ and finding f_i Alice knows that Bob' state is $|\phi_i\rangle$ however Bob does not, unless Alice tells him (we will say they *communicate classically*) the result of her measurement.

Local measurements can have a non-local effect. If Alice performs a local measurement on an entangled state, the system will collapse to a separable one, and Bob's state will depend on the result of Alice's measurement.

For the moment let us forget about Bob and let us try and assign a density matrix $\hat{\rho}_A$ to describe only Alice system and accommodate for this lack of (classical) knowledge regarding Bob, this will introduce the concept of *Reduced Density Matrix*.

5.5 Reduced Density Matrix

LECTURE 13

Given a bipartite (or multi-partite) system, we can define the partial trace over a subsystem to be a trace in that subsystem only. I.e. the partial trace over system B (or Hilbert space $\mathcal{H}_B$) is defined by

$$\mathrm{Tr}_B\left(\hat{C} \otimes \hat{D}\right) = \mathrm{Tr}(\hat{D})\,\hat{C}$$

and all other properties follow from linearity. Note that the partial trace Tr_B maps linear operators acting on $\mathcal{H}_A \otimes \mathcal{H}_B$ to linear operators acting only on $\mathcal{H}_A$, i.e. the result of a partial trace is NOT a number but rather an operator on the remaining Hilbert space. Similarly the partial trace over A

$$\mathrm{Tr}_A\left(\hat{C} \otimes \hat{D}\right) = \mathrm{Tr}(\hat{C})\,\hat{D}\,,$$

produces an operator on $\mathcal{H}_B$.

We define the *reduced density matrix* for a subsystem to be the partial trace of the density matrix over the other subsystem(s). I.e. for a bipartite system

$$\hat{\rho}_A \equiv \mathrm{Tr}_B(\hat{\rho})$$

and

$$\hat{\rho}_B \equiv \mathrm{Tr}_A(\hat{\rho})\,.$$

Generically the reduced density matrices will describe mixed states even if the full system is in a pure state. This is reminiscent of our discussion regarding mixed states: if we "forget" about Bob system, i.e. if we consider the partial trace over B, Alice will have some lack of classical knowledge about her system which means that her state will not be a pure one but rather a mixed state, i.e. the mixed state described by the reduced density matrix.

Example:

Let us consider the two-qubit system $\mathcal{H} = \mathcal{H}_q \otimes \mathcal{H}_q$ in the pure, entangled state

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle \right) .$$

(This will be the prototypical example of entangled states called a Bell state or EPR pair) Firstly let us compute the density matrix in operator form

$$\hat{\rho} = |\Psi\rangle \langle\Psi| = \frac{1}{2} \left(|0\rangle \langle 0| \otimes |0\rangle \langle 0| + |0\rangle \langle 1| \otimes |0\rangle \langle 1| + |1\rangle \langle 0| \otimes |1\rangle \langle 0| + |1\rangle \langle 1| \otimes |1\rangle \langle 1| \right) .$$

We can compute the reduced density

$$\begin{aligned}\hat{\rho}_A = \mathrm{Tr}_B \hat{\rho} = \frac{1}{2} \Big[& |0\rangle \langle 0| \, \mathrm{Tr}_B \left(|0\rangle \langle 0| \right) + |0\rangle \langle 1| \, \mathrm{Tr}_B \left(|0\rangle \langle 1| \right) + \\ & + |1\rangle \langle 0| \, \mathrm{Tr}_B \left(|1\rangle \langle 0| \right) + |1\rangle \langle 1| \, \mathrm{Tr}_B \left(|1\rangle \langle 1| \right) \Big] \\ = \frac{1}{2} & \left(|0\rangle \langle 0| + |1\rangle \langle 1| \right) = \frac{\hat{I}}{2} .\end{aligned}$$

We will see that these Bell states will be maximally entangled, for the moment we have just seen that the reduced density matrix over the second (or first) qubit produces the most mixed qubit state, i.e. half of the identity, i.e. the centre of the Bloch sphere.

We can also obtain this result by first constructing the density matrix as a 4×4 matrix using the standard basis for the usual orthonormal basis states $\{|00\rangle , |01\rangle , |10\rangle , |11\rangle\}$. We first rewrite the state $|\beta_{00}\rangle$ as the 4-dimensional vector

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle \right) \to \mathbf{v} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} ,$$

then as always the density matrix becomes

$$\hat{\rho} = |\beta_{00}\rangle \langle\beta_{00}| \to \rho = \mathbf{v} \, \mathbf{v}^\dagger = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} ,$$

which we rewrite in the sum of tensor operators as

$$\rho = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} ,$$

and finally perform the partial trace over the second operator

$$\begin{aligned}\rho_A = & \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \mathrm{Tr} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \mathrm{Tr} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \\ & + \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \mathrm{Tr} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \mathrm{Tr} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\ = & \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \frac{\mathbb{I}_2}{2} ,\end{aligned}$$

as already obtained above.

The reduced density matrix $\hat{\rho}_A$ is a density matrix which describes Alice's subsystem when she has no knowledge of Bob's system. The following properties illustrate this:

- $\hat{\rho}_A$ is invariant under all LO in system B. For unitary transformations this is a simple consequence of cyclicity of the trace. For measurements in B this is true provided Alice does not know the results of any measurements. This explains why Alice typically sees a mixed state (before she performs any measurements.)
- Under unitary transformations in system A, $\hat{\rho}_A$ transforms as expected for a density matrix.
- Local measurements in system A can be described in terms of operators acting on $\mathcal{H}_A$ and $\hat{\rho}_A$.

For the point about measurement, the main result is that the probability and final state can be calculated using $\hat{F}_i$ and $\hat{\rho}$, or using $\hat{F}_{Ai}$ and $\hat{\rho}_A$, and either method gives the same predictions and results. Specifically

$$\mathrm{Tr}_B\left(\hat{F}_i\,\hat{\rho}\,\hat{F}_i\right) = \hat{F}_{Ai}\,\hat{\rho}_A\,\hat{F}_{Ai}.$$

The reduced density matrix captures a lot of information regarding the nature of the state in consideration. In particular we have the following theorem.

Theorem: If the system $\mathcal{H}_A \otimes \mathcal{H}_B$ is in a pure state $|\Psi\rangle$ then the reduced density matrix $\hat{\rho}_A = \mathrm{Tr}_B\hat{\rho}$ is pure if and only if $|\Psi\rangle$ is separable, i.e. $|\Psi\rangle = |\psi\rangle \otimes |\phi\rangle$ with $|\psi\rangle \in \mathcal{H}_A$ and $|\phi\rangle \in \mathcal{H}_B$.

Proof. ($\Leftarrow$) Let us start with the separable pure state $|\Psi\rangle = |\psi\rangle \otimes |\phi\rangle$. Then $\hat{\rho} = |\Psi\rangle\langle\Psi| = |\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|$, so for the reduced density matrix we have

$$\begin{aligned}\hat{\rho}_A &= \mathrm{Tr}_B\left(\hat{\rho}\right) = \mathrm{Tr}_B\left(|\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|\right)\\ &= |\psi\rangle\langle\psi|\,\mathrm{Tr}_B\left(|\phi\rangle\langle\phi|\right) = |\psi\rangle\langle\psi|\ ,\end{aligned} \tag{5.1}$$

since the state $|\phi\rangle$ is normalised, i.e. $\mathrm{Tr}_B\left(|\phi\rangle\langle\phi|\right) = \langle\phi\,|\phi\rangle = 1$. So if the starting pure state $\hat{\rho}$ is separable then the reduced density matrix is pure $\hat{\rho}_A = |\psi\rangle\langle\psi|$.

($\Rightarrow$) Conversely let us assume the reduced density matrix $\hat{\rho}_A = |\psi\rangle\langle\psi|$ is given by a pure state. Let us complete the vector $|\psi\rangle$ to an orthonormal basis for $\mathcal{H}_A$, i.e. $\mathcal{H}_A = \mathrm{span}\{|\psi\rangle, |\psi_i^\perp\rangle\}$ with $\langle\psi\,|\psi_i^\perp\rangle = 0$ and $\left\langle\psi_i^\perp\,\middle|\psi_j^\perp\right\rangle = \delta_{ij}$.

The most general state in $\mathcal{H}_A \otimes \mathcal{H}_B$ can be written as

$$|\Psi\rangle = c\,|\psi\rangle \otimes |\phi\rangle + \sum_i c_i\left|\psi_i^\perp\right\rangle \otimes |\phi_i\rangle\ ,$$

for some $|\phi\rangle, |\phi_i\rangle \in \mathcal{H}_B$ and $c, c_i \in \mathbb{C}$.

If Alice measures the observable $\left|\psi_j^\perp\right\rangle\left\langle\psi_j^\perp\right| \otimes \hat{I}$ on $|\Psi\rangle$ she has the expectation value

$$\mathrm{Tr}_{A\otimes B}\left[|\Psi\rangle\langle\Psi|\left(\left|\psi_j^\perp\right\rangle\left\langle\psi_j^\perp\right| \otimes \hat{I}\right)\right] = |c_j|^2\,,$$

using the orthonormality properties of the vectors $|\psi_i^\perp\rangle$. But this must be equal to

$$\mathrm{Tr}_A\left(\hat{\rho}_A\left|\psi_j^\perp\right\rangle\left\langle\psi_j^\perp\right|\right)=\left\langle\psi\left|\psi_j^\perp\right\rangle\left\langle\psi_j^\perp\right|\psi\right\rangle=0\,,$$

hence $c_j=0$ for all j which means that the state $|\Psi\rangle=c\,|\psi\rangle\otimes|\phi\rangle$ is indeed separable.

□

Measurement destroys entanglement.

Corollary: If the spectrum of $\hat{F}_A$ is non-degenerate then measuring $\hat{F}_A$ in the system $\mathcal{H}_A$ produces a separable state on the system $\mathcal{H}_A\otimes\mathcal{H}_B$. In other words, measurement destroys entanglement.

Proof. We know that if we measure $\hat{F}_A$ on the state $\hat{\rho}_A$ and we find the outcome f_i we must collapse the wave function to the one dimensional (because the spectrum is non-degenerate) eigenspace spanned by the corresponding vector $|i\rangle$. Hence we have that the reduced density matrix $\hat{\rho}_A$ must go to

$$\hat{\rho}_A\rightarrow\hat{\rho}'_A=|i\rangle\langle i|\ ,$$

but since the new density matrix $\hat{\rho}'_A$ is clearly pure we must have from the theorem above that the state $\hat{\rho}$ in $\mathcal{H}_A\otimes\mathcal{H}_B$ has collapsed to a separable state

$$\hat{\rho}\rightarrow\hat{\rho}'_A\otimes\hat{\rho}'_B=|i\rangle\langle i|\otimes|\phi_i\rangle\langle\phi_i|\ .$$

□

With this new concept of reduced density matrix we can also understand why entanglement, although non-local in nature, does not violate causality. Let us start again with our favourite entangled pure state

$$\begin{aligned}|\Psi\rangle&=\sum_{i,m}\gamma_{im}\,|i\rangle\otimes|m\rangle\ ,\\&=\sum_i\alpha_i\,|i\rangle\otimes|\phi_i\rangle\ ,\\&=\sum_n\beta_m\,|\psi_m\rangle\otimes|m\rangle\ ,\end{aligned}$$

where the various coefficients have been defined as above.

We know that Alice measuring $\hat{F}=\hat{F}_A\otimes\hat{I}$ will collapse the state $\hat{\rho}\rightarrow\hat{\rho}'_A\otimes\hat{\rho}'_B$. If Bob could detect this we would have that Alice measurement would result in instantaneous communication which violates the causality of physics, i.e. any signal should travel from Alice to Bob no faster than the speed of light.

However how could Bob detect this? Well the only thing he can do is perform a measurement for his favourite observable $\hat{G} = \hat{I} \otimes \hat{G}_B$, so what we are really trying to understand is whether the probabilities of outcomes for Bob have changed before and after Alice's measurement.

Before Alice measures we know that Bob has the reduced density matrix

$$\hat{\rho}_B = \text{Tr}_A \left(|\Psi\rangle \langle\Psi| \right) = \sum_{i,j} \alpha_i \alpha_j^* |\phi_i\rangle \langle\phi_j| \, \text{Tr}_A \left(|i\rangle \langle j| \right) = \sum_i |\alpha_i|^2 |\phi_i\rangle \langle\phi_i| \ ,$$

i.e. Bob has the ensemble $\{(|\alpha_i|^2, |\phi_i\rangle\}$.

Now Alice measure $\hat{F}$ and finds outcome f_i so she knows that $|\Psi\rangle \to |i\rangle \otimes |\phi_i\rangle$ which means

$$\hat{\rho} \to \hat{\rho}' = |i\rangle \langle i| \otimes |\phi_i\rangle \langle\phi_i| = \hat{\rho}'_A \otimes \hat{\rho}'_B$$

where the now collapsed density matrix $\hat{\rho}'_B$ is now different from the reduced density matrix $\hat{\rho}_B$ compute above, i.e. $\hat{\rho}'_B = |\phi_i\rangle \langle\phi_i| \neq \hat{\rho}_B$.

However Bob does not know that Alice has measured and found f_i unless Alice communicates this information, Bob only knows that he has the state $|\phi_i\rangle$ with probability $|\alpha_i|^2$. Although Alice has instantaneously changed Bob system, Bob cannot detect this, there has not been any instantaneous transmission of information between Alice and Bob. Said differently if Alice and Bob have prepared 100 copies of the same state $|\Psi\rangle$ and for a hundred times Alice has measure $\hat{F}$, unless she classically communicates in which instances she has found outcome f_i, Bob cannot detect any difference in his probability distributions because he cannot possibly know which one are the instances for which Alice has found outcome f_i or outcome say f_1!

To summarise: Bob *cannot*

- Instantaneously detect the result of Alice's measurement: we have already seen that the order of measurements is irrelevant, the probability of outcome (f_i, g_m) is $|\gamma_{im}|^2$;
- Know what Alice has measured: Suppose that Alice chooses a different observable $\hat{F}'$ this will just select a different orthonormal basis $\{|\tilde{i}\rangle\}$ of eigenstates of $\hat{F}'$ but still we have that

  $$|\Psi\rangle = \sum_i \alpha_i |i\rangle \otimes |\phi_i\rangle = \sum_i \tilde{\alpha}_i |\tilde{i}\rangle \otimes \left|\tilde{\phi}_i\right\rangle \ ,$$

  for some new normalised coefficients $\tilde{\alpha}_i$ and states $\left|\tilde{\phi}_i\right\rangle \in \mathcal{H}_B$. So Bob's ensemble can be thought as $\{(\,|\alpha_i|^2, |\phi_i\rangle)\}$ or $\{(\,|\tilde{\alpha}_i|^2, \left|\tilde{\phi}_i\right\rangle)\}$ but still the mixed state is described by the same reduced density matrix $\hat{\rho}_B = \text{Tr}_A \hat{\rho}$.
- Even know that Alice has made a measurement: otherwise this would mean that Bob's result could depend on whether or not Alice made a measurement at any point in time, even in the future! Clearly not possible in quantum mechanics.

However so far we have assumed that Alice and Bob do not communicate at all. The non-locality of quantum mechanics cannot be detected with local operations only (LO) but as soon as we add classical communications (CC) between Alice and Bob everything changes. If Alice calls Bob to tell him the result of her measurements Bob can indeed detect the change in his state and the story gets interesting.

5.6 Classical Communication

If we allow Alice and Bob to communicate by classical means, i.e. sending classical bits, then some interesting possibilities arise. Note that classical communication now travels at finite speed (at best the speed of light, or a very fast pigeon) so we do not have any problem with causality. The key point is that this allows them to take actions on their own systems based on information provided by the other.

If this is only information which could have been communicated in advance, it is not adding anything new – i.e. we normally assume Alice and Bob have pre-arranged any procedures to follow and that they know the initial state of the full system. So, the only new thing to communicate would be the result of a measurement. The effect of this is that if Bob informs Alice about a measurement he made, this could give Alice some information about her system – this will happen provided the state before the measurement was entangled. Alice can then decide what to do based on this information. If Alice and Bob can act with LO and Classical Communication (CC) we say that they can use LOCC.

Sometimes we allow arbitrary classical communication. Other times we want to consider questions such as how many bits are required to convey enough information to carry out a specific process. In applications where we consider issues of security, we assume that any communications can be intercepted (by Eve) without detection by Alice or Bob. Of course, if say Alice send some bits to Bob, Eve can copy the bits being transmitted before forwarding them to Bob, and so she can know everything about the transmissions. She could of course do other things such as modify the data being transmitted or not send on anything to Bob.

5.7 Quantum Communication

This means that we allow Alice and Bob to send quantum states to each other. If we allow arbitrary quantum communication, we don't really have a bipartite system. E.g. Bob could send his whole system to Alice. She could then perform arbitrary operations on the complete system, and then send Bob's part back to him. Obviously there is no sense in which the two parts of the system were separated or non-interacting. Instead, if we allow quantum communication at all we typically impose specific restrictions, such as Alice can send one qubit to Bob. In this way we can explore questions such as what could be done with a single qubit compared to a single classical bit of communication.

In applications concerned with security, again we assume Eve can intercept any transmissions. However, unlike classical communications, Eve cannot make a copy of the qubits (due to the no-cloning theorem). Also, unless she already knows that they are eigenstates of a specific measurement operator, she cannot measure them without some non-zero probability that she disturbs the state through the measurement process. This means that if Eve gains any information about the qubits before forwarding them, this can be detected with non-zero probability. The result is that quantum communication can be used to ensure security in a way which is not possible with classical communication. See specifically the discussion of Quantum Key Distribution.

5.7.1 Hardware

Although in this module we will entirely focus on the "software" part, to better understand quantum communication is perhaps useful to see the "hardware" of quantum computers, i.e. the physical realization of quantum computers, or even just two-level systems, i.e. the qubit.

In a real world scenario we have three important quantities: the decoherence time τ_Q, i.e. the time it takes the environment to corrupt our quantum system, the operation time τ_{op}, i.e. the time it takes to perform unitary transformations, and maximum number of operations $n_{op} = \tau_Q/\tau_{op}$, roughly how many operations we can do to our system before it is destroyed.

I will not present here the pros and cons for various physical realisation but if you are interested Chapter 7 of Nielsen and Chuang discusses all these topics.

Perhaps the simplest apparatus are optical photons. A single photon state can be produced in a lab by attenuating the output of a laser, furthermore we can act on photons using mirrors, phase shifters, beamsplitters and can be made interacting using nonlinear optical media. This to say that we can indeed produce various states, entangled or not, and act on them with unitary transformation. We can take an electromagnetic cavity (think about two very good mirrors at distance λ) and produce a quantum superposition of zero and one photon of wave-length λ bouncing back and forth, i.e. a qubit $a\,|0\rangle + b\,|1\rangle$. Alice and Bob can then produce an entangled state and each one of them store their qubit (photon/no-photon superposition) in two separate cavities. Alice will then perform some unitary transformation on her qubit and then (very carefully) give the cavity to Bob who has now access to the whole two-qubit system and can perform his favourite 4×4 dimensional unitary transformation.

Another candidate for real world qubits are single-atom cavities. We can think of a single atom standing between two mirrors (called a Fabry-Perot cavity). For simplicity let us assume that this atom energy levels are just two, the lowest called ground state with energy $E_0 = 0$, and an excited state with energy $E_1 > E_0$. We can then have a single photon with energy $E_2 - E_1$ interacting with this single atom since it's resonant, i.e. the atom in the ground state can "eat" the photon and go to the excited state. Quantum information can then be treated in various ways: for example we can use photon states (quantum superposition of 0 and 1 photons) and the cavity with atom provides the non-linear interactions between them, or represented by the atom (quantum superposition of ground and excited state) with photons communicating between different atoms.

The final example uses the spin of particles, being that the orbitals of atoms (ion traps) or the nuclear spin states (NMR the same as the medical device!). The key point is that now we can use magnetic fields to interact with spin systems. For example we can start with a spin-1/2 particle, say for example an electron (although these are not the particles used in real systems) that can be described by specifying the component of its spin by convention along the z-axis, hence our qubit is $a\left|\uparrow\right\rangle + b\left|\downarrow\right\rangle$ where $\left|\uparrow\right\rangle$ would be a state rotating counter-clockwise along the z-axis and $\left|\downarrow\right\rangle$ clockwise. Couplings between different electrons (or chemical bonds between neighbouring atoms the real world realisation) provide the interactions to produce entangled states. So the only Alice and Bob can share two entangled states, each one of them act with some magnetic field on their on spin system and then send it to the other person.

Recently Google has announced that their new quantum processor: "Sycamore" had reached quantum supremacy. They have a two dimensional array of 54 of what they are called *transmons* qubits where each qubit is coupled to the four nearest neighbours. Their qubits are obtained from superconducting circuits for which the conduction electrons condense into Cooper pairs (a macroscopic quantum effect), these two superconducting islands are coupled via two Josephson junctions which are just two superconducting regions separated by a barrier. These Cooper pairs are pseudo-particles formed by two electrons paired together and they are the fundamental objects in the theory of superconductors. The qubit is now realised by the quantum superposition of a Cooper pair transferred between these two islands.

6

Entanglement applications

LECTURE 14

We focus on bipartite systems which are in pure states. From a QI point of view separable states are not interesting. If we have a separable pure state, we simply have two separate quantum systems, each with their own pure state. There is no quantum correlation between such systems so at most there are classical correlations. However, there are interesting effects when we have an entangled state. Then it turns out to be possible to do things which could not be done otherwise. Some examples we explore are:

- Teleportation – The transmission of a quantum state using classical communication, but no quantum communication.
- Quantum Key Distribution – The ability to share a secret random key, using classical and quantum communication, with no possibility that the key could be know by Eve.
- Superdense coding – The ability to transmit 2 classical bits of information by sending just one qubit.

Note that entanglement implies a quantum correlation which is different from classical correlations. In particular even though we have a bipartite system, measurement in one subsystem can instantaneously affect the other. It is this “spooky action at a distance” which Einstein objected to.

However, to reiterate our discussion from above: note that this instantaneous effect does not imply faster than light communication. The main point is that say if Alice makes a measurement, although this can alter Bob’s state the measurement process is random, so she cannot choose how Bob’s state is altered. Furthermore, Bob cannot determine which measurement Alice made, or even if she made a measurement at all. This is because Bob’s reduced density matrix is not changed by any unitary transformations of measurements made by Alice, it describes the same mixed state. If Alice tells Bob the result of a measurement, Bob can then learn something about his state which would change his mixed state (perhaps to a pure state). However, this process requires classical communication so there cannot be any transmission of information faster than the speed of light. Another way to say this is that if Alice and Bob both make a measurement, the probability distribution of their combined results does not depend on whether Alice measures first, Bob measure first or if they measure simultaneously.

A simple example of classical correlation is the following. Charlie gives Alice and Bob each a box with a coin in it. One coin is heads up, the other tails, and Alice and Bob both know this. Alice can open her box to find out whether her coin is heads or tails, and she would then know immediately that Bob’s coin was the other, whether or not Bob had opened his box. Obviously the result is independent of whether or not Bob already opened his box, and also obviously Alice's measurement has no effect on Bob's coin.

A similar example of quantum correlation would be that Charlie gives Alice and Bob each a box containing a 'quantum coin', again with the guarantee that if they both measure to heads/tails, one of them will finds heads, the other tails. Now the difference to the classical case is that before measuring, neither coin is heads or tails, measuring changes the state to either heads or tails at random. Then if Alice measures heads, Bob's coin will instantaneously become tails (assuming Bob has not yet made a measurement.) Similarly if Bob measures first, his measurement will affect Alice's coin and her coin will instantly be heads if he found tails. However, in QM we do not have any mechanism to describe a signal travelling from one system to the other. Indeed, the measurements could be made simultaneously and they will still yield opposite results, yet we cannot claim that one affected the other[1].

Now, a natural question is, can we really distinguish this quantum correlation from classical correlation? More generally, how can we be sure that the state is not either heads or tails before we measure? For any given process it is easy to come up with possible ways that the measurement result is pre-determined, i.e. where our uncertainty about the result of a measurement is simply due to our lack of knowledge about some "hidden variables".

It sounds very difficult to argue against this possibility, but in fact Bell derived an inequality that must hold in any theory with the property of *local realism*. This means any theory where the result of any measurement is pre-determined (i.e. is determined by the state and does not involve any randomness) and local in the sense that no event can affect any other event unless some sort of signal travels (no faster than the speed of light) to communicate the first event to the second. It is easy to show that QM can violate Bell inequalities, and experiment has confirmed that they are violated in nature. This proves that no theory obeying local realism (i.e. no hidden variable theory) can be the correct description of nature. (This does not prove QM is correct, but QM is consistent with all experiments, and there is no known alternative.)

6.1 Bell States

Many of the features of entanglement can be explored in the simplest bipartite system, where each subsystem is a single qubit. The complete system is then a 2-qubit system, so the Hilbert space has dimension 4. We can use an orthonormal basis of separable states

$$\Big\{ |x\rangle \otimes |y\rangle \ : \ x, y \in \{0,1\} \Big\} .$$

In general, linear combinations of these states will be entangled. Given a specific state, one way to check if it is separable is simply to try and write it in the form $|\psi\rangle \otimes |\phi\rangle = a\,|0\rangle \otimes |\phi\rangle + b\,|1\rangle \otimes |\phi\rangle$ for some $a, b \in \mathbb{C}$ and some state $|\phi\rangle$. If this is not possible, the state is entangled. Another way is to check if the reduced density matrix (of either subsystem) gives a pure state, meaning that the state is separable,

[1] In Special Relativity the concept here is spacelike separation – there is no possibility for any signal at or below the speed of light to travel from Alice to Bob or Bob to Alice to allow the result of one measurement to influence the other. In fact in relativity there is no concept of which measurement was made before the other if they were made at a spacelike separation.

or a mixed state, meaning that the state is entangled. Recall that the state is pure iff. $\mathrm{Tr}(\rho^2) = \mathrm{Tr}(\rho) = 1$.

In fact $\mathrm{Tr}(\rho_A^2) = \mathrm{Tr}(\rho_B^2)$ gives a measure of the entanglement, with maximum value 1 for no entanglement (i.e. a separable state) to the minimum value $1/2$ (for a single qubit subsystem) for maximally entangled states.

For maximally entangled states, the trace of the reduced density matrix of both sides is $1/2$.

The four Bell states

$$|\beta_{xy}\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle \otimes |y\rangle + (-1)^x \, |1\rangle \otimes |\overline{y}\rangle \Big)$$

where $\overline{y} = NOT\ y$ (defined by $\overline{0} = 1$ and $\overline{1} = 0$) are maximally entangled, and also form an orthonormal basis.

Note that in terms of a 2-qubit system the Bell state basis is related to the standard basis by a unitary transformation, so these are entirely equivalent choices of basis states. However, for the bipartite system the Bell states cannot be created from the separable states by any LOCC process. This is because the required unitary transformations are not of the form $\hat{U}_A \otimes \hat{U}_B$ as that would only transform a separable state to a separable state. Also, measurement by Alice and/or Bob would also result in a separable state. Note that a measurement by Alice or Bob on a Bell state can result in a separable state such as $|0\rangle \otimes |0\rangle$ but of course measurement is not a reversible transformation.

On the other hand, it is possible for Alice or Bob to individually transform any Bell state to any other Bell state. Specifically, the unitary operators $\hat{U}_{xy} \otimes \hat{I}$ (which Alice can use) and $\hat{I} \otimes \hat{U}_{xy}$ (which Bob can use) transform the Bell state $|\beta_{00}\rangle$ to the Bell state $|\beta_{xy}\rangle$, i.e.

$$\hat{U}_{xy} \otimes \hat{I}\, |\beta_{00}\rangle = \hat{I} \otimes \hat{U}_{xy}\, |\beta_{00}\rangle = |\beta_{xy}\rangle \ ,$$

where the unitary operators are given in the standard single qubit matrix representation by

$$\begin{aligned} U_{00} = \mathbb{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad &, \quad U_{01} = \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \\ U_{10} = \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad &, \quad U_{11} = i\sigma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \ . \end{aligned}$$

Note that when dealing with qubit, or multi-qubits systems we will often drop the hat from operators and use, unless differently specified, the basis $|0\rangle\,,|1\rangle$ represented as the two standard basis vectors $(1,0)^T\,,(0,1)^T$, hence $\hat{I}$ for a qubit will become $\mathbb{I}_2$ the 2×2 identity matrix, and in this basis we will directly use the Pauli matrices instead of having to define three abstract operators (usually called spin operators $\hat{S}_x\,,\hat{S}_y\,,\hat{S}_z$ or sometimes $\hat{J}_x\,,\hat{J}_y\,,\hat{J}_z$) represented by these three matrices in this given basis.

> **Example:**
>
> Consider $\hat{U}_{11} = i\sigma_2$, then we have
>
> $$\hat{U}_{11}\,|0\rangle = -\,|1\rangle \ , \qquad \hat{U}_{11}\,|1\rangle = |0\rangle \ .$$

> Then if we consider
>
> $$\hat{U}_{11} \otimes \hat{I} \left|\beta_{00}\right\rangle = \hat{U}_{11} \otimes \hat{I} \left(\frac{1}{\sqrt{2}} \left|0\right\rangle \otimes \left|0\right\rangle + \frac{1}{\sqrt{2}} \left|1\right\rangle \otimes \left|1\right\rangle \right)$$
> $$= \left(-\frac{1}{\sqrt{2}} \left|1\right\rangle \otimes \left|0\right\rangle + \frac{1}{\sqrt{2}} \left|0\right\rangle \otimes \left|1\right\rangle \right) = \left|\beta_{11}\right\rangle \ ,$$
>
> as expected.

6.2 Superdense Coding

This process is called superdense because we can use one qubit to transmit two bits of information. However, this is only possible at the cost of using a resource of entanglement corresponding to one Bell state. I.e. we cannot convey more than one bit of information using a single qubit unless the two parties already share an entangled state.

To see this, first note that it is obvious that a qubit can be used in place of a bit. E.g. simply send the state $\left|0\right\rangle$ corresponding to the bit 0, or the state $\left|1\right\rangle$ corresponding to the bit 1. The recipient just measures, corresponding to the operator $\frac{1}{2}\left(\mathbb{I}_2 - \sigma_3\right)$ to distinguish, with probability 1, these two orthogonal states

$$\frac{1}{2}\left(\mathbb{I}_2 - \sigma_3\right)\left|0\right\rangle = 0\left|0\right\rangle \ , \qquad \frac{1}{2}\left(\mathbb{I}_2 - \sigma_3\right)\left|1\right\rangle = 1\left|1\right\rangle \ ,$$

i.e. the outcome of the measurement is with probability 1 the classical bit Alice wanted to send.

Note as well that Alice can indeed prepare the initial qubit to represent in this way the classical bit she wants to send to Bob. Suppose she wants to send the classical bit $x \in \{0, 1\}$, she can just prepare a random state $\left|\psi\right\rangle$ and measure on it $\frac{1}{2}\left(\mathbb{I}_2 - \sigma_3\right)$. The result of this measurement can only be 0 or 1, if it is precisely the bit x she wants to send then she knows she has now projected $\left|\psi\right\rangle$ to the right state $\left|x\right\rangle$ and she sends it straight away.

If the result of the measurement is not the same as the bit she wanted to send then she knows she has projected the state onto $\left|\bar{x}\right\rangle$ and she just needs to perform the unitary evolution $\sigma_1\left|\bar{x}\right\rangle = \left|x\right\rangle$ and then she can send the state to Bob.

Now, although a qubit contains two real numbers (as it corresponds to a point on a sphere, as we saw in the Bloch sphere picture), the recipient can only make one measurement which will give one of two possible values, and then the state will be transformed to one of two orthogonal states. (Which states depend on the choice of measurement operator.)

Now, suppose Alice and Bob share the Bell state $\left|\beta_{00}\right\rangle$ and Alice wants to send the 2-digit binary number $(xy)_2$ to Bob. She can do this by transforming the system by acting on her qubit with the unitary operator $\hat{U}_{xy}$ which transforms the whole system to the state $\left|\beta_{xy}\right\rangle$. Note that this does not transmit any information to Bob, his reduced density matrix is $\hat{\rho}_B = \frac{1}{2}\hat{I}$ before and after Alice's transformation.

Superdense coding relies on the fact that Alice can transform $\left|\beta_{00}\right\rangle$ to any $\left|\beta_{xy}\right\rangle$ using only a local unitary transformation on her part of the Hilbert space.

Alice can then send her qubit to Bob. Note that this qubit also contains no information on its own. This means that if say Eve intercepts it, she just has a qubit with reduced density matrix $\frac{1}{2}\hat{I}$ completely independent of x and y. However, assuming

Bob receives the qubit from Alice, he will have the full Bell state $|\beta_{xy}\rangle$. Since the four Bell states are orthogonal, he can distinguish them (with probability 1) with a suitable measurement. This means any measurement operator which has the four Bell states as eigenstates (with distinct eigenvalues.) E.g. Bob will definitely measure the result $(xy)_2$ if he measures

$$\hat{B} = 0\,|\beta_{00}\rangle\,\langle\beta_{00}| + 1\,|\beta_{01}\rangle\,\langle\beta_{01}| + 2\,|\beta_{10}\rangle\,\langle\beta_{10}| + 3\,|\beta_{11}\rangle\,\langle\beta_{11}|\ .$$

6.3 No-Cloning Theorem

Crucial to the claim above that one qubit could only be used to transmit one bit of information (without use of entanglement) was that fact that the recipient could only perform one measurement without destroying the original state. From our knowledge of measurement, this is obvious. However, suppose Bob received a qubit and then cloned it. He could then measure on the many copies to deduce (to any desired accuracy by making more copies) the probability distribution of the results of any measurement(s). E.g. doing this for the measurement $\frac{1}{2}(\mathbb{I}_2 - \sigma_3)$ would give an estimate of $\cos^2(\theta/2)$ (defined on the Bloch sphere) as the probability of getting the result corresponding to $|0\rangle$ rather that $|1\rangle$. This would give an arbitrary number of bits of information as the binary representation of this number, for example $\cos^2(\theta/2) = (0.1100010101...)_2$. However, the *No-Cloning Theorem* states that this is not possible.

Theorem: In QM it is impossible to clone an unknown state $|\psi\rangle$. I.e. we cannot transform $|\psi\rangle \otimes |\Omega\rangle \to |\psi\rangle \otimes |\psi\rangle$ for arbitrary unknown $|\psi\rangle$, where $|\Omega\rangle$ is a fixed initial state.

Note, the attempt here is to create a quantum photocopier, taking any input $|\psi\rangle$ and a blank sheet of paper $|\Omega\rangle$ and making a perfect copy. Note that it is possible to transform $|\psi\rangle \otimes |\Omega\rangle \to |\phi\rangle \otimes |\psi\rangle$ for arbitrary $|\psi\rangle$ but the state $|\phi\rangle$ will not (in general) be the state $|\psi\rangle$, or even depend on the state $|\psi\rangle$.

Proof. Note that measurement cannot help. Any measurement of an unknown state will give a result dependent on the choice of measurement, and the final state will be an eigenstate of the measurement operator. We can only deduce that the original state was not orthogonal to this final state. This means that such a quantum copier must use unitary evolution. However, we can prove by contradiction that there is no such unitary transformation. Alternatively we can prove that there is no such linear transformation. (Either is sufficient to prove the theorem.)

Linearity: Take any two linearly independent states $|\psi_1\rangle$ and $|\psi_2\rangle$ and assume we have a quantum copier. Also, let $|\psi\rangle = a\,|\psi_1\rangle + b\,|\psi_2\rangle$ for any non-zero $a, b \in \mathbb{C}$. Then the copier acts as

$$\begin{aligned} |\psi_1\rangle \otimes |\Omega\rangle &\to |\psi_1\rangle \otimes |\psi_1\rangle \\ |\psi_2\rangle \otimes |\Omega\rangle &\to |\psi_2\rangle \otimes |\psi_2\rangle \\ |\psi\rangle \otimes |\Omega\rangle &\to |\psi\rangle \otimes |\psi\rangle \end{aligned}$$

but by linearity we must also have

$$|\psi\rangle \otimes |\Omega\rangle = a\,|\psi_1\rangle \otimes |\Omega\rangle + b\,|\psi_2\rangle \otimes |\Omega\rangle \to a\,|\psi_1\rangle \otimes |\psi_1\rangle + b\,|\psi_2\rangle \otimes |\psi_2\rangle \neq |\psi\rangle \otimes |\psi\rangle\ .$$

Hence we have a contradiction.

Unitarity: The argument is similar. Take all states to be normalised and consider two states $|\psi_1\rangle$ and $|\psi_2\rangle$. The copier must act as

$$\begin{aligned} |\psi_1\rangle \otimes |\Omega\rangle &\rightarrow |\psi_1\rangle \otimes |\psi_1\rangle \\ |\psi_2\rangle \otimes |\Omega\rangle &\rightarrow |\psi_2\rangle \otimes |\psi_2\rangle \end{aligned}$$

but since it is a unitary operation, inner products are preserved. This means that

$$(\langle\psi_1| \otimes \langle\Omega|)\,(|\psi_2\rangle \otimes |\Omega\rangle) = (\langle\psi_1| \otimes \langle\psi_1|)\,(|\psi_2\rangle \otimes |\psi_2\rangle)$$

but this means that

$$\langle\psi_1\,|\psi_2\rangle = (\langle\psi_1\,|\psi_2\rangle)^2 \ .$$

This is only possible if $\langle\psi_1\,|\psi_2\rangle = 1$ so $|\psi_1\rangle = |\psi_2\rangle$ or $\langle\psi_1\,|\psi_2\rangle = 0$ so the states are orthogonal. So, again we see it is impossible to copy arbitrary unknown states. □

Note that the above does *not* mean that we cannot clone an arbitrary state $|\psi\rangle$. It is just that we must choose the right unitary transformation, depending on $|\psi\rangle$ and obviously we cannot do that if we don't know anything about the state.

> **Example:**
>
> It *is* possible to write down a unitary operator which transforms
>
> $$U\big(\,|n\rangle \otimes |0\rangle\,\big) \rightarrow |n\rangle \otimes |n\rangle \ . \tag{6.1}$$
>
> In the matrix notation, this operator is given by
>
> $$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} . \tag{6.2}$$
>
> But despite the fact that (6.1) looks like a generic copier, the states on which it acts are *not* the most general one-qubit states. After all, the most general such state is
>
> $$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \ . \tag{6.3}$$
>
> If you act with the U given above on this state $|\psi\rangle$, you will find (try it!) that it does not produce $|\psi\rangle \otimes |\psi\rangle$. So the copier only works for the basis states, not for linear combinations of them.

6.4 Teleportation

Teleportation is the process of transferring a quantum state, say from Alice to Bob, without any quantum communication. Surprisingly, although it is not possible to clone, it is still possible to teleport an unknown quantum state using only classical communication, although this does require the resource of entanglement. It is in some sense complementary to the process of superdense coding, where entanglement is again the resource. In both cases two bits correspond to one qubit, and we require one pair of maximally entangled qubits as the resource to convert in either direction.

Consider the simplest example where Alice has one unknown qubit state $|\psi\rangle = a|0\rangle + b|1\rangle$. Note that even if she knew what the state was, in general she would need to communicate two real numbers (e.g. the two angles giving the position of the Bloch sphere) to Bob so he could create a copy. Obviously, to any reasonable accuracy, this would require many bits of information, and certainly just 2 bits would not be enough. However, this is not what happens in teleportation.

In the process below, Alice and Bob do not need to know anything about the state $|\psi\rangle$, they do not learn anything about it in the teleportation process, and Bob does not create a copy as at the end Alice no longer has the state $|\psi\rangle$.

The starting point is that Alice and Bob must share an entangled state. Assume they share the Bell state $|\beta_{00}\rangle$ so the state of the whole system is

$$|\psi\rangle \otimes |\beta_{00}\rangle = \frac{1}{\sqrt{2}}|\psi\rangle \otimes |0\rangle \otimes |0\rangle + \frac{1}{\sqrt{2}}|\psi\rangle \otimes |1\rangle \otimes |1\rangle$$

where Alice has the first two qubits, and Bob has the third.

Now Alice and Bob can use LOCC to teleport the state $|\psi\rangle$, i.e. so that the final state of the system is $|\Phi\rangle \otimes |\psi\rangle$, and Alice's final 2-qubit state $|\Phi\rangle$ does not depend on $|\psi\rangle$. Alice then no longer has the state $|\psi\rangle$, but Bob does – it has been teleported.

The process is as follows.

- Alice entangles $|\psi\rangle = a|0\rangle + b|1\rangle$ with the other two qubits. She can do this with a unitary transformation on her two qubits since the second is already entangled with Bob's qubit.

  First we act with the *controlled-NOT (CNOT) gate* given by the unitary operator transforming

  $$|00\rangle \to |00\rangle \ , \quad |01\rangle \to |01\rangle \ , \quad |10\rangle \to |11\rangle \ , \quad |11\rangle \to |10\rangle$$

  on Alice's two qubits producing the state

  $$\begin{aligned} |\psi\rangle \otimes |\beta_{00}\rangle &= \frac{1}{\sqrt{2}}\Big(a|000\rangle + a|011\rangle + b|100\rangle + b|111\rangle\Big) \\ \text{CNOT}&\Big\downarrow \\ \hat{U}_{\text{CNOT}} \otimes \hat{I}\,|\psi\rangle \otimes |\beta_{00}\rangle &= \frac{1}{\sqrt{2}}\Big(a|000\rangle + a|011\rangle + b|110\rangle + b|101\rangle\Big) \ . \end{aligned}$$

Passing to the standard basis representation the CNOT gate is given by the 4×4 matrix

$$U_{\text{CNOT}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \left(\frac{\mathbb{I}_2 + \sigma_3}{2}\right) \otimes \mathbb{I}_2 + \left(\frac{\mathbb{I}_2 - \sigma_3}{2}\right) \otimes \sigma_1 \,, \tag{6.4}$$

where we note that U_{CNOT} is not of the form $A \otimes B$ (see discussion in section 5.3) which is the reason why the unitary operator $U_{\text{CNOT}} \otimes \mathbb{I}_2$ entangles Alice two qubits and leaves Bob qubit invariant.

- If she then measures her two qubits, Bob's qubit state will depend on the result of her measurement. This can be done so that Bob's qubit is a unitary transformation of the state $|\psi\rangle$ that we want to teleport. (To clarify this, we present the measurement process as a unitary transformation followed by measurement to distinguish the standard basis states. Instead Alice could simply measure to distinguish the states $|\pm\rangle \otimes |y\rangle$.)

  First Alice acts on the first qubit with the unitary operator mapping $|0\rangle \to |+\rangle$ and $|1\rangle \to |-\rangle$ called the *Hadamard gate*. This is just given by the unitary operator (change of basis matrix)

$$U_H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \,,$$

  which clearly implements

$$\hat{U}_H |0\rangle = |+\rangle \;, \qquad \hat{U}_H |1\rangle = |-\rangle \;.$$

  So after this unitary evolution we have

$$\begin{aligned} &\hat{U}_H \otimes \hat{I} \otimes \hat{I} \frac{1}{\sqrt{2}}\Big(a\,|000\rangle + a\,|011\rangle + b\,|110\rangle + b\,|101\rangle\Big) = \\ &= \frac{1}{2}\Big[a\,(|0\rangle + |1\rangle) \otimes (|00\rangle + |11\rangle) + b\,(|0\rangle - |1\rangle) \otimes (|10\rangle + |01\rangle)\Big] = \\ &= \frac{1}{2}\Big[\,|00\rangle \otimes (a\,|0\rangle + b\,|1\rangle) + |01\rangle \otimes (a\,|1\rangle + b\,|0\rangle) + \\ &\qquad + |10\rangle \otimes (a\,|0\rangle - b\,|1\rangle) + |11\rangle \otimes (a\,|1\rangle - b\,|0\rangle)\Big] \,, \end{aligned}$$

  where we used all linearities properties of tensor products.

  We can rewrite this expression as the state

$$\frac{1}{2}\sum_{x,y} |x\rangle \otimes |y\rangle \otimes \hat{U}_{xy}\,|\psi\rangle \;,$$

  where $\hat{U}_{xy}$ are the unitary transformations defined above when we introduced the Bell states

$$\hat{U}_{00} = \mathbb{I}_2 \,, \qquad \hat{U}_{01} = \sigma_1 \,, \qquad \hat{U}_{10} = \sigma_3 \,, \qquad \hat{U}_{11} = i\sigma_2 \,.$$

  Alice will now measure her two qubits (e.g. measuring $(\mathbb{I}_2 - \sigma_3)/2$ for each of them) and she will get one of the four results $(xy)_2$ with $x, y \in \{0, 1\}$, which means that Bob's state will then be $\hat{U}_{xy}\,|\psi\rangle$. After Alice measures $(\mathbb{I}_2 - \sigma_3)/2$ for both her qubits with outcomes x, y entanglement is destroyed and we are left with the separable state $|x\rangle \otimes |y\rangle \otimes \hat{U}_{xy}\,|\psi\rangle$.

- Bob can then recover the state $|\psi\rangle$ by a unitary transformation of his qubit, but which one depends on the result of Alice's measurement so she must communicate which of the four possible results she got – this requires 2 bits of classical communication. Therefore Alice sends the two bits of information giving the values of x and y, and Bob then acts on his state $\hat{U}_{xy}|\psi\rangle$ with $\hat{U}_{xy}^{-1} = \hat{U}_{xy}^{\dagger}$ so his state is then $|\psi\rangle$. I.e. we have teleported the state $|\psi\rangle$ from Alice lab to Bob's at the price of using the Bell pair shared by Alice and Bob.

  Note that we have not cloned $|\psi\rangle$! Alice does not have this state any longer, she is left with the two qubits $|x\rangle \otimes |y\rangle$, which are independent of $|\psi\rangle$. Note also that neither Alice nor Bob need to know what the teleported state $|\psi\rangle$ is at all and with this protocol they have not gained any additional information about it.

Both with teleportation and superdense coding we have achieved something classically impossible, however, in both protocols we have expended some resources: for teleportation we have destroyed the Bell state by performing a measurement thus destroying entanglement and producing a separable state, while in superdense coding Alice had to give away her qubit and the Bell pair is not shared anymore.

6.5 Quantum Key Distribution

LECTURE 16

This is an application using QM to ensure that Alice and Bob can produce a secret shared random key. Such a key can then be used for absolutely secure communication. Specifically, QKD allows them to generate a random string of bits so that they will both know the value of each bit, and can be sure no one else does. Of course, they could do this without any quantum theory by meeting but this may not be convenient and they would have to know in advance how long the key should be, and they would have to keep the key secure until they needed to use it. Instead, QKD allows them to do this using quantum communication

Note that the purpose is secure classical communication, but QKD itself does not communicate any information. Instead, QKD allows Alice and Bob to share a key which Alice can use to encrypt a message. The encrypted message can then be transmitted (using classical communication) to Bob. He can then use the same key to decrypt the message. The point is that even if the transmission is intercepted, the encrypted message does not provide any information about the actual message without knowledge of the key.

To see how this works consider a message M represented as a binary number (note that this can be done in any other basis) of n bits and a random secret key K also n bits long shared by Alice and Bob. Since the encryption works independently on each bit, consider a bit m of the message and the corresponding bit k in the key. We do not make any assumptions about the value of m, but we assume that k has values 0 and 1 with equal probability (and the value is known only by Alice and Bob.) Then the mechanism is as follows:

- Alice produces the encrypted message $C = M \oplus K$ where $\oplus$ means bitwise addition modulo 2, i.e. for each bit $c = m \oplus k$ which is just $m + k$ except $1 \oplus 1 = 0$. This is also sometimes noted as the XOR operation (or exclusive OR): $(True\, XOR\, False) = (False\, XOR\, True) = True$ while $(False\, XOR\, False) =$

$(True\,XOR\;True) = False$.

- Alice transmits the encrypted message C to Bob. If Eve intercepts this, she gains no information about M since for each bit c there is equal probability that this is the same as m or different from m. Since there are only two possible values for each bit, C is completely random and independent of M without knowledge of the key K.

- Bob decodes the message since by calculating $C \oplus K$ since by modular arithmetic
$$C \oplus K = (M \oplus K) \oplus K = M \oplus (K \oplus K) = M \oplus 0 = M \ .$$

Note that it is vital that the key is as long as the message and it is not reused. E.g. if the same bit k is used for two bits m_1 and m_2 of the message the property that either $m_1 = m_2$ or that $m_1 \neq m_2$ is unchanged for c_1 and c_2. A key shorter than the message would make the encryption susceptible to attacks.

Note that for a key as long as the message this encryption method is not susceptible to any sort of attack. One cannot apply any frequency analysis because all the characters are equally likely.

However this method has an obvious drawback! Requiring such a new key is a significant practical obstacle to using this encryption method. This is why this method was called *One-Time Pad* (OTP) as the key can be used only once and then it has to be destroyed. Furthermore the key cannot be transmitted in clear and required to be physically transported (by a spy wearing a fake mustache carrying an inconspicuous briefcase handcuffed to their wrist) from the sender to the recipient, or the be agreed before-hand, for example by using a pre decided book or the daily newspaper.

Standard encryption methods (such as the ones you use daily on the internet) instead rely on *trapdoor functions*. These are functions where even with full knowledge of the function f itself, it is very difficult to find the inverse function f^{-1}, but if you know f^{-1} it is easy to find f. In this case Bob can choose such a function f^{-1} and then tell Alice what f is. She then sends $C = f(M)$ and Bob decrypts it using $f^{-1}(C) = M$. This is very convenient since there is no need to share a secret key. The disadvantage is that it is possible to find f^{-1} with knowledge of f. Normally this is fine since either the information in the message is of no use after a certain time, or we assume that it is not worth the significant effort to find it. On the other hand, in some instances it may be important that the information is never found by a third party, and for such applications trapdoor functions are not secure.

So the main problem of OTP is the transmission of the encryption key, how does QM help exchange a secret random key without Alice and Bob having to meet? There are many different protocols, but one is the *BB84 Protocol*:

- Alice makes a random choice 0 or 1 (with equal probability) and also a random choice X or Z. According to her choices (which she records), she prepares a single qubit state
$$(0, Z) \to |0\rangle \ , \ \ (1, Z) \to |1\rangle \ , \ \ (0, X) \to |+\rangle \ , \ \ (1, X) \to |-\rangle \ .$$
She sends this qubit to Bob (using quantum communication).

- Bob receives the qubit and randomly chooses Z, in which case he measures $\frac{1}{2}(\mathbb{1}_2 - \sigma_3)$, or X, in which case he measures $\frac{1}{2}(\mathbb{1}_2 - \sigma_1)$ (remember what these operators mean on the Bloch sphere).

- Alice and Bob repeat this process as often as required to generate a sufficiently long key.

- Alice and Bob announce (publicly, so no secrecy is assumed) their choices of X or Z for each qubit and discard all results where they did not make the same choice. Where they made the same choice, Bob's measurement is guaranteed to agree with Alice's choice of 0 or 1. This generates the random shared key. (If it's not long enough, they can repeat to extend the key.)

Why is this secure? Suppose Eve intercepts a qubit. What can she do before forwarding it to Bob? To learn something she must measure. However, since the four possible states are not all orthogonal, Eve cannot make a measurement which will distinguish them with certainty. E.g. suppose Eve chose to measure $\frac{1}{2}(\mathbb{1}_2 - \sigma_3)$. If Alice had chosen Z, Eve would get the result 0 or 1 matching Alice's key, and Eve would forward the qubit unchanged to Bob. However, with equal probability Alice would have chosen X, in which case for either $|\pm\rangle$, Eve would get a random result 0 or 1 and forward correspondingly a random qubit $|0\rangle$ or $|1\rangle$ to Bob. If Bob chose to measure X this result would be discarded (as Alice chose Z) but if Bob chose Z, he would get the random result of Eve's measurement, so half the time this would differ from Alice's key.

Alice choice	**Eve Choice**	**Bob Choice**	**Results**
Z	Z	Z	A and B and E measures all match: BAD
Z	Z	X	Discarded
Z	X	X	Discarded
Z	X	Z	In 50% of this case E measures match A. In 50% of this case B measures does *not* match A.

Table 6.1: List of possible cases. Similar if Alice chooses X.

The above means that overall, for each bit of the shared key which Eve has intercepted and measured, there is a 25% chance that Alice and Bob will not have the same value for their keys as schematically depicted in Table **??**. If Alice and Bob do nothing, Eve will know 75% of the key which could be disastrous. However, if Alice and Bob compare a random subset of their keys, they can estimate the error rate. If it is too high, they will assume interference and discard the key, hoping to repeat the process. If they get a low enough error rate, they can assume the worst case scenario that all the errors are due to Eve. However, they can reduce the key in a way which makes it highly unlikely Eve will have any information about it.

Note that there are many other ways to exchange a key. E.g. if Alice and Bob share Bell states such as $|\beta_{00}\rangle$ they could simply both measure $\frac{1}{2}(\mathbb{1}_2 - \sigma_3)$, getting the same random result. This is sufficient if they already share enough entangled qubits. However, if Alice prepares the Bell states and sends one qubit to Bob, Eve could simply also measure $\frac{1}{2}(\mathbb{1}_2 - \sigma_3)$ then forward the state to Bob. Eve would then learn the full key without being detected. To circumvent this, Alice and Bob can employ the same strategy as above, as provided they both choose X or both choose Z, they will get the same results. Again Eve will be detected with probability 25% for each bit. Note that here, and above, it is essential that Alice and Bob do not announce any choice of X or Z before they are sure Bob has received the qubit.

6.6 Bell Inequalities

LECTURE 17 & 18

Einstein Podolsky and Rosen were not particularly happy with the non-local ("spooky action at a distance") and probabilistic ("God does not play dice") nature of quantum mechanics. So they postulated that quantum mechanics must be incomplete! Any "complete" theory must satisfy the postulate of *local realism*, i.e.

- Locality: no faster than light influences;
- Realism: measurements must be deterministic, i.e. the results tell us a property of the system.

This lead to the notion of *hidden variables* theories. According to EPR there must be some extra parameters not included in quantum mechanics that if we were to measure would give us a fully deterministic world. Quantum mechanics then only seems probabilistic because we lack the knowledge about these hidden variables.

However Bell showed that any such theory with local realism must satisfy a certain type of inequality while quantum mechanics violate this! There are different inequalities which are all generally called *Bell Inequalities*. We consider a specific version known as the CHSH (Clauser, Horne, Shimony & Holt) Inequality.

YouTube
Short Qiskit video on the Bell inequality published after the 2022 Nobel Prize was given to Clauser, Aspect and Zeilinger for their related work.

In all cases the feature is that a result is derived which must be obeyed by any theory satisfying certain conditions such as *locality* (an effect at one point can be detected at another point only if something travels between these points) and some definition of *realism* meaning that measurements always reveal a property of the system (i.e. the system really had that property whether we measured or not.)

Roughly speaking, classical physics should obey these conditions whereas QM does not seem to. The question is to make this precise, and Bell inequalities allow us to demonstrate that QM does not satisfy these conditions, and so no classical *hidden variable* theory can reproduce all predictions of QM. The huge advantage of this approach is that we don't have to rule out candidate hidden variable theories one by one.

6.6.1 CHSH Bell-Inequality

Suppose we have a system with four observables Q, R, S and T which each can take only the values ± 1. The realism property tells us that any state of this system must have specific values for these four observables, i.e. (q, r, s, t) .

Take a large number of states of this system (which can have different values of these observables) and measure the quantity $QS + RS + QT - RT$ for each such state and calculate the average, i.e. the expectation value $E(QS + RS + QT - RT)$. Now, due to the restricted values for each observable $Q = \pm R$, so on each state when we measure we either have $Q + R = 0$ and $Q - R = \pm 2$ or $Q + R = \pm 2$ and $Q - R = 0$. Hence either $(Q + R)S = 0$ and $(Q - R)T = \pm 2$ or $(Q + R)S = \pm 2$ and $(Q - R)T = 0$, so $QS + RS + QT - RT$ can only take the values ± 2. Obviously taking the average we must have

$$-2 \leq E(QS + RS + QT - RT) = \\ E(QS) + E(RS) + E(QT) - E(RT) \leq 2\ . \quad (6.5)$$

Note that the above argument used realism since we assumed that each state of this system had definite values of all 4 observables, i.e. we can really assign values to both Q and R for each state.

Now, consider the following *EPR experiment*: we have a system where Alice and Bob are separated (as far as we want) and Charlie is exactly in the middle.

1. Charlie will prepare lots and lots of Bell states $|\beta_{11}\rangle$ and send one qubit of each simultaneously to Alice and Bob so that they receive them at exactly the same time.
2. On receiving each qubit Alice will make a random choice of Q or R and immediately measure the qubit, and similarly Bob will measure randomly either S or T. Assuming locality, this setup excludes the possibility that Alice or Bob's measurement can affect the other via something travelling at finite speed[2].
3. If our quantum mechanics were really a local realism theory we would have that Alice and Bob results are predetermined by some hidden variable that describe the Bell state sent out by Charlie.
4. For each qubit Alice and Bob record their choice of measurement and the result. This way they can later compare and calculate $E(QS)$, $E(RS)$, $E(QT)$ and $E(RT)$.

Note that each Bell state is only contributing to one of those expectation values, so Alice and Bob never really measure $QS + RS + QT - RT$ for any individual state. However, by the assumption of realism, measuring say $E(QS)$ on a random subset of states will give a good estimate (subject only to statistical errors) of $E(QS)$ averaged over all states.

Specifically, we can take measurements using matrices (in the standard representation)

$$Q = \sigma_1 \otimes \mathbb{1}_2\,, \qquad R = \sigma_3 \otimes \mathbb{1}_2$$

which Alice can measure using only her qubit, and

$$S = \mathbb{1}_2 \otimes \frac{-1}{\sqrt{2}}(\sigma_1 + \sigma_3)\,, \qquad T = \mathbb{1}_2 \otimes \frac{-1}{\sqrt{2}}(\sigma_1 - \sigma_3)$$

which Bob can measure.

Note that since $[Q, R] \neq 0$ and $[S, T] \neq 0$ we cannot simultaneously measure both Q and R, or similarly S and T. Hower both Alice and Bob only measure one and then we just consider the average over many copies of the same state.

[2] Knowing the time taken to measure, the distance between Alice and Bob give a lower bound on the speed of any propagation. It is certainly possible to ensure that any such propagation must be faster than the speed of light, and so violate relativity.

Using QM we predict that the outcome of each measurement is either ± 1 (Check the eigenvalues of all these operators) and that for the Bell state $|\beta_{11}\rangle$ we can calculate

$$\begin{aligned}
E(QS) = \langle QS \rangle = \langle \beta_{11} | QS | \beta_{11} \rangle = \langle \beta_{11} | (\sigma_1 \otimes \mathbb{I}_2) \left(\mathbb{I}_2 \otimes \frac{-1}{\sqrt{2}} (\sigma_1 + \sigma_3) \right) |\beta_{11}\rangle \\
&= -\frac{1}{2\sqrt{2}} \left(\langle 01| - \langle 10| \right) \sigma_1 \otimes (\sigma_1 + \sigma_3) \left(|01\rangle - |10\rangle \right) \\
&= -\frac{1}{2\sqrt{2}} \left(\langle 01| - \langle 10| \right) \mathbb{I}_2 \otimes (\sigma_1 + \sigma_3) \left(|11\rangle - |00\rangle \right) \\
&= -\frac{1}{2\sqrt{2}} \left(\langle 01| - \langle 10| \right) \left(|10\rangle - |01\rangle - |11\rangle - |00\rangle \right) \\
&= -\frac{1}{2\sqrt{2}} (-1 - 1) = \frac{1}{\sqrt{2}} ,
\end{aligned}$$

and similarly (exercise)

$$E(RS) = E(QT) = -E(RT) = \frac{1}{\sqrt{2}} .$$

However, the combination

$$E(QS) + E(RS) + E(QT) - E(RT) = \frac{4}{\sqrt{2}} = 2\sqrt{2} > 2$$

so this clearly violates the Bell (CHSH) inequality.

Experiments have confirmed this violation of the Bell inequality, so nature is not described by a theory obeying local realism. Nature is consistent with QM.

However there are loopholes to this argument. In particular there is something called super-determinism where everything is predetermined, including the "randomness" of Alice and Bob choices of the observable to measure. There is no such thing as free will and all our choices and actions are dictated by a fully deterministic world. Whenever one wants to understand the foundations of quantum mechanics inevitably the discussion will turn philosophical and for us it is a good point to stop.

7

Information theory

We review some basic concepts of classical and quantum information theory. Essentially this is about quantifying measures of information and quantum entanglement. Typically when we quantify information we describe it in terms of a number of bits, corresponding to the minimum length message required to convey that information. However, note that typically the length of a message will be much longer than this as most communication systems contain redundancy. E.g. a paragraph with correct punctuation and grammar will be longer than a text message, yet both can convey exactly the same information, and even most text messages could be compressed (but at the cost of readability.)

7.1 Classical Information and Shannon Entropy

LECTURE 19

One way to describe information is through *Shannon Entropy* which gives the average number of bits required to specify a message from a set of possible messages where we know the probability of each message. In terms of probability theory, we consider a random variable X and define $p(x)$ to be the probability that $X = x$. Then the Shannon entropy $H(X)$ is defined to be

$$H(X) = -\sum_x p(x) \log p(x)$$

where conventionally we take $\log$ to mean logarithm base 2. We also take the convention that $0 \log 0 = 0$ as we don't expect very small probabilities to be much different from probability zero, and $\lim_{p\to 0} p \log p = 0$.

This definition can be interpreted as either a measure of how unsure we are about the value of X, or on average how much information we gain when we learn the value of X. Note that the average is over all possible values x, weighted by the probability that $X = x$. The quantity $-\log p(x)$ is a choice of measure of information. The interpretation is that for $p(x) \simeq 1$, we do not gain much information if X takes that value since we anyway expected that to be the case. However, if we find $X = x$ for some value with small $p(x)$, that is surprising and we gain a lot of information. The choice of normalisation, or reason for taking logarithms base 2, is so that if x can take one of 2^N values with equal probability, $H(X) = N$. It is fairly obvious that we could not encode such message using fewer than n bits. The fact that in general $H(X)$ is the lower bound on the average number of bits required to encode the messages is *Shannon's Noiseless Coding Theorem*. Also note that if we have a fixed set of possible message, $H(X)$ is maximised when the probability distribution is uniform. A standard example of coding which attains the bound is:

> **Example:**
>
> Suppose there are four possible values with $p(1) = 1/2$, $p(2) = 1/4$ and $p(3) = p(4) = 1/8$. Then $H(X) = 7/4$. Obviously we could encode the four possible messages using two bits per message. However, we can achieve $7/4$ bits on average with the following coding, using shorter message for more common messages:
>
> $$1 \to 0\ ,\quad 2 \to 10\ ,\quad 3 \to 110\ ,\quad 4 \to 111\ .$$
>
> It is easy to check that the average length (weighted by the above probabilities) is $7/4$. Note that the coding is such that we don't need any extra bits to indicate the start or finish of the message. The receiver knows that 0 or the third consecutive 1 indicates the end of the message. Such messages can be sent one after the other without any ambiguity.

Now consider entropies involving two random variables X and Y.

7.1.1 Joint Entropy

This is written

$$H(X,Y) = -\sum_{x,y} p(x,y)\log p(x,y)\ .$$

It obeys a property called *subadditivity*:

$$H(X,Y) \le H(X) + H(Y)\ .$$

It is easy to see that $H(X,Y) = H(X) + H(Y)$ when X and Y are independent variables, i.e. when $p(x,y) = p(X=x)p(Y=y)$.

7.1.2 Relative Entropy

This is defined for two random variables which take the same values but with different probability distributions, say $p(x)$ and $q(x)$. The relative entropy of the distribution $p(x)$ to $q(x)$ is

$$\begin{aligned} H(p(x)||q(x)) = \sum_x \left(p(x)\log p(x) - p(x)\log q(x)\right) \\ = -H(X) - \sum_x p(x)\log q(x)\ . \end{aligned} \tag{7.1}$$

The relative entropy is non-negative and

$$H(p(x)||q(x)) = 0 \iff p(x) = q(x) \ \ \forall x\ .$$

A corollary of this is the subadditivity property of joint entropy, with equality if and only if the variables are independent. To see this, calculate the relative entropy of the distribution of variables X and Y with probabilities $p(x,y)$ to the distribution with probabilities $p(X=x)p(Y=y)$ where $p(X=x) = \sum_y p(x,y)$ and $p(Y=y) = \sum_x p(x,y)$.

7.1.3 Conditional Entropy and Mutual Information

The *conditional entropy* of X given Y is the average (over the values of Y) uncertainty left about X when we know Y, or equivalently the information we gain by learning the value of X if we already knew the value of Y. It is given by

$$H(X|Y) = H(X,Y) - H(Y) \le H(X)\ .$$

A related concept is the *mutual information* of X and Y which is the amount of information we gain about one by knowing the other, or equivalently the information shared (or duplicated) by X and Y rather than being in one of them alone:

$$H(Y:X) = H(X:Y) = H(X) + H(Y) - H(X,Y) \\ = H(X) - H(X|Y) = H(Y) - H(Y|X) \geq 0 \,. \quad (7.2)$$

LECTURE 20

7.2 Quantum Entropy

The *von Neumann entropy* of a quantum state with density operator $\hat{\rho}$ is defined to be

$$S(\hat{\rho}) = -\text{Tr}\left(\hat{\rho}\log\hat{\rho}\right) \,.$$

By interpreting the state as an ensemble (of one or more) orthogonal pure states, this is just the Shannon entropy of that ensemble. I.e. we can always diagonalise the density matrix to write $\hat{\rho} = \sum_i p_i \left|i\right\rangle\left\langle i\right|$ where $\left|i\right\rangle$ are orthonormal which we can use to form (at least part of) a basis. Then the density matrix is a diagonal matrix with entries p_i (and perhaps some zeroes) where these diagonal elements are the eigenvalues of $\hat{\rho}$. Then

$$-\text{Tr}\left(\hat{\rho}\log\hat{\rho}\right) = -\sum_i p_i \log p_i = H(p_i) \,.$$

Note that for a pure state $S(\hat{\rho}) = -1\log 1 = 0$.

As for the classical case, we can use this to define various related concepts.

7.2.1 Relative Entropy

This is a measure of the difference (or in fact a notion of the distance) between two states (in the full system) with density matrices $\hat{\rho}_1$ and $\hat{\rho}_2$:

$$S(\hat{\rho}_1||\hat{\rho}_2) = \text{Tr}\left(\hat{\rho}_1\log\hat{\rho}_1\right) - \text{Tr}\left(\hat{\rho}_1\log\hat{\rho}_2\right) \geq 0$$

with equality if and only if $\hat{\rho}_1 = \hat{\rho}_2$.

7.2.2 Joint Entropy, Conditional Entropy and Mutual Information

If we have a bipartite system with Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, a density matrix $\hat{\rho}$ and reduced density matrices $\hat{\rho}_A$ and $\hat{\rho}_B$ for each system, we write

$$S(A) = S(\hat{\rho}_A)\,, \quad S(B) = S(\hat{\rho}_B)\,, \quad S(A,B) = S(\hat{\rho})$$

where $S(A,B)$ is called the *joint entropy* of systems A and B.

By analogy with the classical case we can also define the *conditional entropy*

$$S(A|B) = S(A,B) - S(B)$$

but unlike the classical case this can be negative. Indeed, if $\hat{\rho}$ is a pure state, $S(A,B) = 0$ but this does not imply $S(B) = 0$ so we have in this case $S(A|B) = -S(B) \leq 0$. The point here is that initially we know everything about the system since it is in a pure state, but say Bob makes a measurement, this typically results

Measurement turns a pure state (entropy $S(A,B) = 0$) to a mixed state (entropy $S(A) = S(B) > 0$). The conditional entropy $S(A|B)$ can thus be *negative*.

in Alice having a mixed state so she can only be certain what her state is when Bob communicates the result of the measurement.

The mutual information is written $I(A:B)$ or $S(A:B)$ and is given by

$$I(A:B) = S(A) + S(B) - S(A,B) = S(A) - S(A|B) = S(B) - S(B|A) \ .$$

In the case where $\hat{\rho}$ is a pure state

$$I(A:B) = S(A) + S(B) - S(A,B) = S(A) + S(B) = 2S(A)$$

so we see that entanglement can be interpreted as mutual information, i.e. the information shared by A and B and not in either subsystem alone. This agrees with what we have seen for Bell states which are maximally entangled 2-qubit states. The reduced density matrix for each qubit is $\frac{1}{2}I$. As we have seen earlier, each qubit alone gives no information about which of the four Bell states we have, this must be stored in the entanglement, i.e. it is a shared quantity.

7.3 Bipartite Entanglement Entropy

The aim is to quantify the quantum entanglement (as opposed to the classical correlations) between two subsystems of a quantum system. We will assume that the full system is in a pure state $\hat{\rho}$. Then we define the entanglement entropy to be

$$S(A) = S(B) \ .$$

We will see below that these quantities are indeed equal since the reduced density matrices have the same non-zero eigenvalues, even if $\mathcal{H}_A$ and $\mathcal{H}_B$ have different dimensions (in which case the number of zero eigenvalues differs, but they don't contribute to the von Neumann entropy.)

Entanglement entropy is the Von Neumann entropy of the reduced density matrix.

7.3.1 Schmidt Decomposition and Schmidt Number

Diagonalise $\hat{\rho}_A$ so that we have non-zero probabilities p_i and orthonormal states $|i\rangle \in \mathcal{H}_A$ with

$$\hat{\rho}_A = \sum_i p_i \left|i\right\rangle \left\langle i\right| \ .$$

Note that the range of i is the number of non-zero eigenvalues of $\hat{\rho}_A$ which may be less than the dimension of $\mathcal{H}_A$. Now, since the full system is in some pure state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, we can write

$$|\Psi\rangle = \sum_i \sqrt{p_i} \left|i\right\rangle \otimes \left|\phi_i\right\rangle \ .$$

Checking that this gives the correct expression for $\hat{\rho}_A$ shows that the $|\phi_i\rangle$ are orthonormal states in $\mathcal{H}_B$, and so the number of non-zero probabilities is also no larger than the dimension of $\mathcal{H}_B$. We call this number of non-zero probabilities the *Schmidt Number* N_S. Note that we can calculate

$$\hat{\rho}_B = \sum_i p_i \left|\phi_i\right\rangle \left\langle \phi_i\right|$$

which shows that $S(A) = S(B)$.

The Schmidt number is a (somewhat crude, as it is integer valued) measure of the entanglement between systems A and B. It satisfied the following properties:

- It is invariant under unitary transformation in A or B.
- Any measurements made locally in A or B cannot increase the Schmidt number.
- The Schmidt number is 1 if and only if $\hat{\rho}_A$ and $\hat{\rho}_B$ are pure states, or equivalently $\hat{\rho}$ is a separable pure state.

We know that for a given number of non-zero probabilities, the Shannon entropy is maximised when the probabilities are equal. So here we have

$$S(A) = S(B) = H(p_i) \leq \log N_S \leq \log D$$

where $D = \min\{\dim \mathcal{H}_A, \dim \mathcal{H}_B\}$.

8

Changelog

2024-01-16 Added FAQ example in the no-cloning section.

9

Bibliography

[1] Paul Benioff. “The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines”. In: *Journal of statistical physics* 22 (1980), pp. 563–591.

[2] David Deutsch. “Quantum theory, the Church–Turing principle and the universal quantum computer”. In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 400.1818 (1985), pp. 97–117.

[3] Richard P. Feynman. “Simulating physics with computers”. In: *Int. J. Theor. Phys.* 21 (1982). Ed. by L. M. Brown, pp. 467–488.

[4] Cleve Moler and Charles Van Loan. “Nineteen dubious ways to compute the exponential of a matrix”. In: *SIAM review* 20.4 (1978), pp. 801–836.