Lecture 0

Introduction to Quantum Information Science

QUANTUM INTERFERENCE

ARTUR EKERT

About complex numbers, called probability amplitudes, that, unlike probabilities, can cancel each other out, leading to quantum interference and qualitatively new ways of processing information.

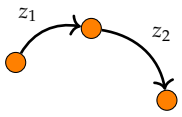
The classical theory of computation usually does not refer to physics. Pioneers such as Alan Turing, Alonzo Church, Emil Post and Kurt Gödel managed to capture the correct classical theory by intuition alone and, as a result, it is often falsely assumed that its foundations are self-evident and purely abstract. They are not! The concepts of information and computation can be properly formulated only in the context of a physical theory – information is stored, transmitted and processed always by *physical* means. Computers are physical objects and computation is a physical process. Indeed, any computation, classical or quantum, can be viewed in terms of physical experiments, which produce *outputs* that depend on initial preparations called *inputs*. Once we abandon the classical view of computation as a purely logical notion independent of the laws of physics it becomes clear that whenever we improve our knowledge about physical reality, we may also gain new means of computation. Thus, from this perspective, it is not very surprising that the discovery of quantum mechanics in particular has changed our understanding of the nature of computation. In order to explain what makes quantum computers so different from their classical counterparts, we begin with the rudiments of quantum theory.

Computation is a physical process!
Computation is a physical process!
Computation is a physical process!
Computation is...
Computation...

1.1. **Two basic rules.** Quantum theory, at least at some instrumental level, can be viewed as a modification of probability theory. We replace positive numbers (probabilities) with complex numbers z (probability amplitudes) such that the squares of their absolute values, $|z|^2$, are interpreted as probabilities. The rules for combining amplitudes are very reminiscent of the rules for combining probabilities:

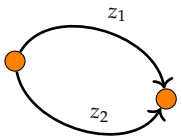
Born's Rule

The correspondence between probability amplitude z and probability $p = |z|^2$ is known as **Born's Rule**.



$z = z_1 z_2$

Whenever something can happen in a sequence of independent steps, we multiply the amplitudes of each step.



$z = z_1 + z_2$

Whenever something can happen in several alternative ways, we add the amplitudes for each separate way.

That's it! These two rules are basically all you need to manipulate amplitudes in any physical process, no matter how complicated. (we will amend the two rules later on when we touch upon the particle statistics). They are universal and apply to any physical system, from elementary particles through atoms and molecules to white dwarfs stars. They also apply to information for, as we have already emphasised, information is physical. The two rules look deceptively simple but, as you will see in a moment, their consequences are anything but trivial.

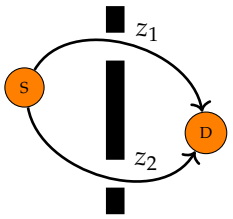
1.2. **Quantum interference and the failure of probability theory.** Modern mathematical probability theory is based on three axioms, proposed by Andrey Nikolaevich Kolmogorov (1903–1987) in his monograph with the impressive German title *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Foundations of Probability Theory). The

Kolmogorov axioms are simple and intuitive. Once you identify all elementary outcomes, or events, you may then assign probabilities to them. Probability is a number between 0 and 1, and an event which is certain has probability 1. These are the first two axioms. There is one more. The probability of any event can be calculated using a deceptively simple rule - the additivity axiom:

> Whenever an event can occur in several mutually exclusive ways, the probability for the event is the sum of the probabilities for each way considered separately.

Obvious, isn't it? So obvious, in fact, that probability theory was accepted as a mathematical framework theory, a language that can be used to describe actual physical phenomena. Physics should be able to identify elementary events and assign numerical probabilities to them. Once this is done we may revert to mathematical formalism of probability theory. The Kolmogorov axioms will take care of the mathematical consistency and will guide us whenever there is a need to calculate probabilities of more complex events. This is a very sensible approach apart from the fact that it does not always work! Today, we know that probability theory, as ubiquitous as it is, fails to describe many common quantum phenomena. In order to see the need for quantum theory let us consider a simple experiment in which probability theory fails to give the right predictions. In a double slit experiment a particle emitted from a source S can reach detector D by taking two different paths, e.g. through an upper or a lower slit. After sufficiently many repetitions of this experiment we can evaluate the frequency of clicks in the detector D and show that it is inconsistent with the predictions based on the probability theory. Let us use the quantum approach to show how the discrepancy arises.

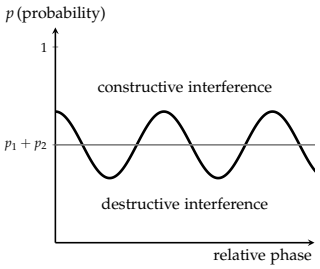
The particle emitted from a source S can reach detector D by taking two different paths, e.g. through an upper or a lower slit, with amplitudes z_1 and z_2 respectively. We may say that the upper slit is taken with probability $p_1 = |z_1|^2$ and the lower slit with probability $p_2 = |z_2|^2$. These are two mutually exclusive events. With the two slits open, probability theory declares (the additivity axiom) that the particle should reach the detector with probability $p_1 + p_2 = |z_1|^2 + |z_2|^2$. Wrong! Following the "quantum rules", first we add the amplitudes and then we square the absolute value of the sum to get the probability. Thus, the particle will reach the detector with probability



$z = z_1 + z_2$

$$\begin{aligned} p = |z|^2 = |z_1 + z_2|^2 &= |z_1|^2 + |z_2|^2 + z_1^\star z_2 + z_1 z_2^\star, \\ &= p_1 + p_2 + |z_1||z_2|(e^{i(\varphi_2 - \varphi_1)} + e^{-i(\varphi_2 - \varphi_1)}), \\ &= p_1 + p_2 + 2\sqrt{p_1 p_2}\ \cos(\varphi_2 - \varphi_1), \\ &= p_1 + p_2 + \text{interference terms}, \end{aligned} \tag{1}$$

where we have expressed the amplitudes in their polar forms $z_1 = |z_1|e^{i\varphi_1}$ and $z_2 = |z_2|e^{i\varphi_2}$. The appearance of the interference terms marks the departure from the classical theory of probability. The probability of any two seemingly mutually exclusive events is the sum of the probabilities of the individual events, $p_1 + p_2$, *modified* by the interference term, $2\sqrt{p_1 p_2}\ \cos(\varphi_2 - \varphi_1)$. Depending on the relative phase $\varphi_2 - \varphi_1$, the interference term can be either negative (destructive interference) or positive (constructive interference), leading to either suppression or enhancement of the total probability p.

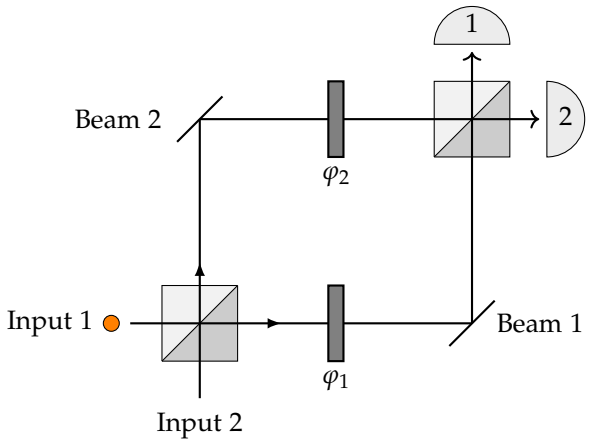


The algebra is simple, our focus is on the physical interpretation. Firstly, note that the important quantity here is the relative phase $\varphi_2 - \varphi_1$ rather than the absolute values φ_1 and φ_2. This observation is not trivial at all. If a particle reacts only to the difference of the two phases, each pertaining to a separate path, then it must have, somehow, experienced the two paths, right? Thus we cannot say that the particle has travelled either through the upper or the lower slit, it has travelled through *both*. In the same way quantum computers follow, in some tangible way, all computational paths simultaneously, producing answers that depend on all these alternative calculations.

Weird, but this is how it is! Secondly, what has happened to the axiom of additivity in probability theory, what is wrong with the additivity axiom? One thing that is wrong is the assumption that the processes of taking the upper or the lower slit are mutually exclusive. In reality, as we have just mentioned, the two transitions *both occur*, simultaneously. However, we cannot learn this from probability theory, or any other a priori mathematical construct. There is no fundamental reason why Nature should conform to the additivity axion. We find out how nature works by making intelligent guesses, running experiments, checking what happens and formulating physical theories. If our guess disagrees with experiments it is wrong, so we try another intelligent guess, and another, etc. Right now quantum theory is the best guess we have; it offers good explanations and predictions that have not been falsified by any of the existing experiments. This said, be assured that one day quantum theory will be falsified and we will have to start guessing again.

According to the philosopher Karl Popper (1902–1994) a theory is genuinely scientific only if it is possible, in principle, to establish that it is false. Genuinely scientific theories are never finally confirmed because no matter how many confirming observations have been made observations that are inconsistent with the empirical predictions of the theory are always possible.

1.2.1. *Example:* One of the simplest quantum devices in which quantum interference can be controlled is a Mach-Zehnder interferometer.



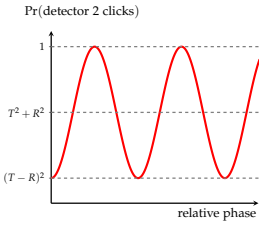
It consists of two beam-splitters (the square boxes, bottom left and top right) and two slivers of glass of different thickness which are inserted into each of the optical paths connecting the two beam-splitters. The slivers are usually referred to as "phase shifters" and their thicknesses, φ_1 and φ_2, are measured in units of the photon's wavelength multiplied by 2π. The two inputs ports of the interferometer are labelled as 1 and 2, and each of the two output ports, also labelled as 1 and 2, terminates in a photodetector. A photon (the orange dot) impinges on the first beam-splitter from one of the two input ports, here input 1, and begins its journey towards one of the two photodetectors. Let U_{ij} denotes the probability amplitude that the photon initially in input port $j = 1, 2$ ends up in detector $i = 1, 2$ (here, and in the following, index i should not be confused with the imaginary unit). At each of the two beam-splitters the photon is transmitted with the probability amplitude $\sqrt{T}$ and reflected with the probability amplitude $i\sqrt{R}$, $(R + T = 1)$, and the two phase shifters modify the amplitudes by phase factors, $e^{i\varphi_1}$ and $e^{i\varphi_2}$, respectively. In quantum theory we almost always start with the amplitudes and once we have a full expression for the amplitude of a given outcome we square its absolute value to get the corresponding probability. For example, let us calculate U_{11}. We notice that there are two alternative ways for the photon in the input port 1 to end up in the output port 1. It can take the lower path, through the phase shifter φ_1, or the upper path, through the phase shifter φ_2. The lower path implies two consecutive transmissions at the beamsplitters and the phase factor $e^{i\varphi_1}$, whereas the upper path implies two consecutive reflections and the phase factor $e^{i\varphi_2}$. Once the photon ends in the output port 1 there is no way of knowing which path was taken, thus we add the amplitudes pertaining to each path.

The resulting amplitude is

$$U_{11} = \sqrt{T}e^{i\varphi_1}\sqrt{T} + i\sqrt{R}e^{i\varphi_2}i\sqrt{R},$$

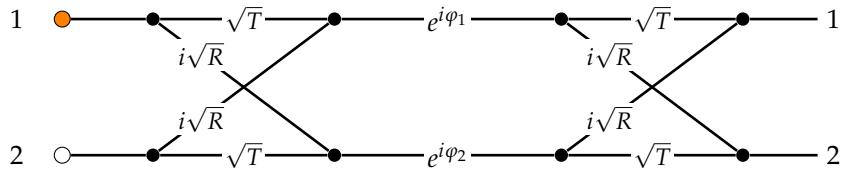
and the corresponding probability $P_{11} = |U_{11}|^2$ reads

$$\begin{aligned} P_{11} &= |\sqrt{T}e^{i\varphi_1}\sqrt{T} + i\sqrt{R}e^{i\varphi_2}i\sqrt{R}|^2 = |Te^{i\varphi_1} - Re^{i\varphi_2}|^2 \\ &= T^2 + R^2 - 2TR\cos(\varphi_2 - \varphi_1). \end{aligned}$$



The "classical" part of this expression, $T^2 + R^2$, basically says that the photon undergoes either two consecutive transmissions with probability T^2, or two consecutive reflections with probability R^2. The probability of being transmitted through any phase shifter is always 1, hence the phase shifters play no role in the classical description of this process. But the classical description is not correct, as the experiments show, and hence the interference term $2TR\cos(\varphi_2 - \varphi_1)$, in which the phase shifters play the essential role. Depending on the *relative* phase $\varphi = \varphi_2 - \varphi_1$ the probability that the detector 1 "clicks" can vary from $(T-R)^2$, for $\varphi = 0$, to 1 for $\varphi = \pi$.

If we do not care about the experimental details, we can represent the action of the Mach-Zehnder interferometer in terms of a diagram:



first beamsplitter phase shifts second beamsplitter

Here, we can follow, from left to right, the multiple different paths that a photon can take in between specific input and output ports. The amplitude for any given path is just the product of the segments, while the overall amplitude is the sum of the amplitudes for the many different paths. You can, for example, see that the probability amplitude U_{21} is given by

$$U_{21} = \sqrt{T}e^{i\varphi_1}i\sqrt{R} + i\sqrt{R}e^{i\varphi_2}\sqrt{T},$$

and the corresponding probability

$$\begin{aligned} P_{21} &= |\sqrt{T}e^{i\varphi_1}i\sqrt{R} + i\sqrt{R}e^{i\varphi_2}\sqrt{T}|^2 \\ &= 2RT + 2RT\cos(\varphi_2 - \varphi_1). \end{aligned}$$

Again, the first term is of "classical" origin and represents probabilities corresponding to each path, one reflection followed by one transmission plus one transmission followed by one reflection, that is, $RT + TR = 2RT$. The second term is the interference term. Clearly, the photon entering port 1 will end up in one of the two detectors, hence,

$$P_{11} + P_{21} = R^2 + 2RT + T^2 = (T+R)^2 = 1.$$

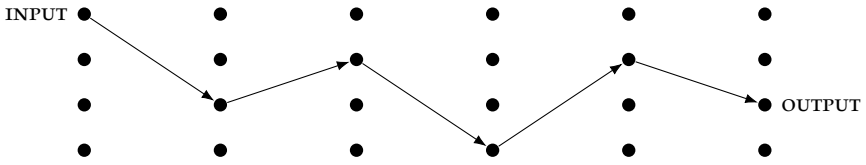
The action of the interferometer is fully described by the four probability amplitudes U_{ij} $(i, j = 1, 2)$. The most popular instance of a Mach-Zehnder interferometer involves only symmetric beamsplitters $(R = T = \frac{1}{2})$ and is fully described by the matrix

$$U = \begin{bmatrix} -\sin\varphi/2 & \cos\varphi/2 \\ \cos\varphi/2 & \sin\varphi/2 \end{bmatrix},$$

In general, any isolated quantum device, including a quantum computer, can be described by a matrix of probability amplitudes U_{ij} that input j generates output i. Watch the order of indices.

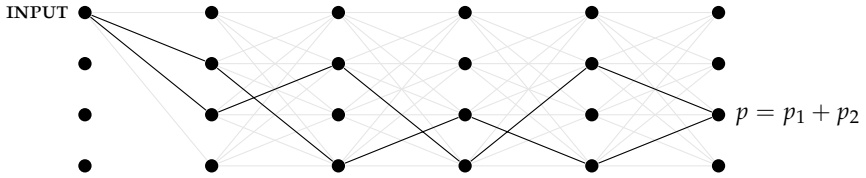
where $\varphi = \varphi_2 - \varphi_1$. In fact, when you do all the calculations you obtain $ie^{i\frac{\varphi_1+\varphi_2}{2}}U$ rather than U, but the global phase factor $ie^{i\frac{\varphi_1+\varphi_2}{2}}$ is common to all the amplitudes in the matrix and as such it does not contribute to the resulting probabilities (why?).

1.3. **Computation.** Think about computation as a physical process that evolves a prescribed initial configuration of a computing machine, called INPUT, into some final configuration, called OUTPUT. We shall refer to the configurations as *states*. The diagram below shows five consecutive computational steps performed on four distinct states.



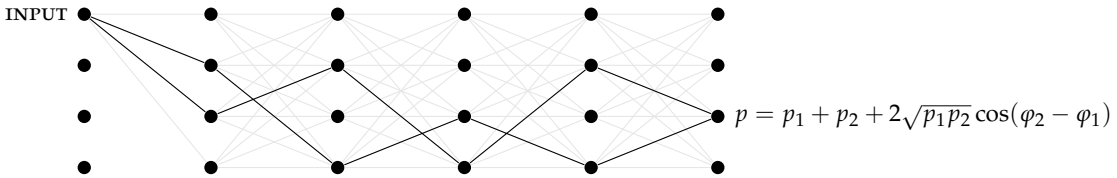
DETERMINISTIC

That computation was *deterministic* – every time you run it with the same input, you get the same output. Such a computation does not have to be deterministic – we can augment a computing machine by allowing it "to toss an unbiased coin" and to choose its steps randomly. It can then be viewed as a directed, tree-like graph where each node corresponds to a state of the machine, and each edge represents one step of the computation.



PROBABILISTIC

The computation starts from some initial state (INPUT) and it subsequently branches into other nodes representing states reachable with non-zero probability from the initial state. The probability of a particular final state (OUTPUT) being reached is equal to the sum of the probabilities along all mutually exclusive paths which connect the initial state with that particular state. The diagram above shows only two computational paths, but, in general, there could be many more of them (here, up to 256) paths contributing to the final probability. Quantum computation can be represented by a similar graph:



QUANTUM

We associate with each edge in the graph the probability *amplitude* that the computation follows that edge. The probability amplitude of a particular path to be followed is the product of amplitudes pertaining to transitions in each step. The probability amplitude of a particular final state being reached is equal to the sum of the amplitudes along all mutually exclusive paths which connect the initial state with that particular state,

$$z = \sum_{\text{all paths k}} z_k.$$

The resulting probability, as we have just seen, is the sum of the probabilities pertaining to each computational path p_k modified by the interference terms,

$$p = |z|^2 = \sum_{k,j} z_j^\star z_k = \sum_k p_k + \sum_{k \neq j} \sqrt{p_k p_j}\cos(\varphi_k - \varphi_j).$$

Quantum computation can be viewed as a complex multiparticle quantum interference involving many computational paths through a computing device. The art of quantum computation is to shape quantum interference, through a sequence of computational steps, enhancing probabilities of correct outputs and suppressing probabilities of the wrong ones.

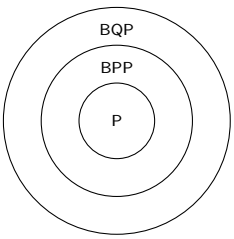
1.4. **Computational Complexity.** Is there a compelling reason why we should care about quantum computation? It may sound like an extravagant way to compute something that can be computed anyway. Indeed, your standard laptop, given enough time and memory, can simulate pretty much any physical process. In principle, it can also simulate any quantum interference and compute everything that quantum computers can compute. The snag is, this simulation, in general, is very inefficient. And efficiency does matter, especially if you have to wait more than the age of the Universe for your laptop to stop and deliver an answer!

The age of the Universe is currently estimated at 13.772 billion years

In order to solve a particular problem, computers (classical or quantum) follow a precise set of instructions — an algorithm. Computer scientists quantify the efficiency of an algorithm according to how rapidly its running time, or the use of memory, increases when it is given ever larger inputs to work on. An algorithm is said to be *efficient* if the number of elementary operations taken to execute it increases no faster than a polynomial function of the size of the input. We take the input size to be the total number of binary digits (bits) needed to specify the input. For example, using the algorithm taught in elementary school, one can multiply two n digit numbers in a time that grows like the number of digits squared, n^2. In contrast, the fastest-known method for the reverse operation—factoring an n-digit integer into prime numbers—takes a time that grows exponentially, roughly as 2^n. That is considered inefficient.

Notice that the technological progress alone, such as increasing the speed of classical computers, will never turn an inefficient algorithm (exponential scaling) into an efficient one (polynomial scaling). Why?

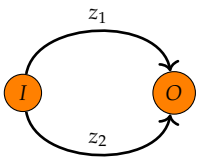
The class of problems that can be solved by a deterministic computer in polynomial time is represented by the capital letter P, for *polynomial* time. The class of problems that can be solved in polynomial time by a probabilistic computer is called BPP, for *bounded-error probabilistic polynomial* time. It is clear that BPP contains P, since a deterministic computation is a special case of a probabilistic computation in which we never consult the source of randomness. When we run a probabilistic, aka randomised, computation many times on the same input, we will not get the same answer every time, but the computation is useful if the probability of getting the right answer is high enough. Finally, the complexity class BQP, for *bounded-error quantum polynomial*, is the class of problems that can be solved in polynomial time by a quantum computer. Since a quantum computer can easily generate random bits and simulate a probabilistic classical computer, BQP certainly contains the class BPP. Here we are interested in problems that are in BQP but not known to be in BPP. The most popular example of such a problem is factoring. A quantum algorithm, discovered by Peter Shor in 1994, can factor n-digit numbers in a number of steps that grows only as n^2. Since the intractability of factorisation underpins the security of many methods of encryption Shor's algorithm was soon hailed as the first 'killer application' for quantum computation, something very useful that only a quantum computer could do. Since then, the hunt has been on for interesting things for quantum computers to do, and at the same time, for the scientific and technological advances that could allow us to build quantum computers.



It must be stressed that not all quantum algorithms are so efficient, in fact many are no faster than their classical counterparts. Which particular problems will lend themselves to quantum speed-ups is an open question.

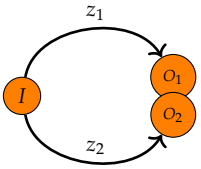
1.5. **Quantum decoherence.** In principle we know how to build quantum computers out of simple components, such as qubits (quantum bits) and quantum logic gates. We will describe these components in detail in the subsequent lectures. However, as the number of quantum components increases, we quickly run into some serious practical problems. The more interacting components are involved, the harder it tends to be to engineer the interactions that would cause the necessary gate operations and induce quantum interference without introducing errors. The more components there are,

the more likely it is that quantum interference will spread outside the quantum computer, to the surrounding environment, thus spoiling the computation. This process is called decoherence. In order to understand the essence of decoherence consider the following two different scenarios in which a quantum computer is prepared in some input state I and generates output O



$p = |z_1 + z_2|^2$

The computer is isolated and quantum computation does not affect the environment. The computer and the environment evolve independently from each other and, as a result, the environment does not hold any physical record of how the computer reached output O. In this case we add the amplitudes for each of the two alternative computational paths.



$p = |z_1|^2 + |z_2|^2$

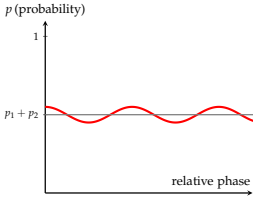
Quantum computation affects the environment. The environment now holds a physical record of how the computer reached output O, which results in two final states of the composed system (computer + environment) which we denote O_1 and O_2. We add the probabilities for each of the two alternative computational paths.

The addition of probability amplitudes, rather than probabilities, applies to physical system which are completely isolated. When quantum computation affects the environment we have to include the environment in our analysis for it now takes part in the computation, i.e. our isolated system is now composed of a quantum computer and its environment. Depending on which computational path was taken the environment may end up in two distinct states. The computer itself may show output O but when we include the environment we have not one but two outputs, O_1 and O_2, denoting, respectively, "computer shows output O and the environment knows that path 1 was taken" and "computer shows output O and the environment knows that path 2 was taken". There are no alternative ways of reaching O_1 or O_2 hence there is no interference and the corresponding probabilities read $p_1 = |z_1|^2$ for O_1, and $p_2 = |z_2|^2$ for O_2. The probability that the computer shows output O, regardless the state of the environment, is the sum of of the two probabilities $p = p_1 + p_2$. We have lost the interference term and with it any advantages of quantum computation. In the presence of decoherence the interference formula Eq.(1) is modified and reads,

$$p = p_1 + p_2 + 2v\sqrt{p_1 p_2}\,\cos(\varphi_2 - \varphi_1),$$

where the parameter v, called the "visibility" of the interference pattern, ranges from 0 (the environment can perfectly distinguish between the two paths, total decoherence, no interference) to 1 (the environment cannot distinguish between the two paths, no decoherence, full interference), with the values in between corresponding to partial decoherence. We shall derive this formula later on and you will see that v quantifies the degree of distinguishability between O_1 and O_2. The more environment knows about which path was taken the less interference we see.

Decoherence suppresses quantum interference.



Decoherence is chiefly responsible for our classical description of the world – without interference terms we may as well add probabilities instead of amplitudes. While decoherence is a serious impediment to building quantum computers, depriving us of the power of quantum interference, it is not all doom and gloom; there are clever ways around decoherence such as the quantum error correction and fault-tolerant methods we will meet later.

1.6. **Outlook.** When the physics of computation was first investigated, starting in the 1960s, one of the main motivations was a fear that quantum-mechanical effects might place fundamental bounds on the accuracy with which physical objects could render the properties of the abstract entities, such as logical variables and operations, that

appear in the theory of computation. But it turned out that quantum mechanics imposes no significant limits but does break through some of those that classical physics imposed. The quantum world has a richness and intricacy that allows new practical technologies, and new kinds of knowledge. In this course we will merely scratch the surface of the rapidly developing field of quantum computation. We will concentrate mostly on the fundamental issues and skip many experimental details. However, it should be mentioned that quantum computing is a serious possibility for future generations of computing devices. At present it is not clear how and when fully-fledged quantum computers will eventually be built; but notwithstanding this, the quantum theory of computation already plays a much more fundamental role in the scheme of things than its classical predecessor did. I believe that anyone who seeks a fundamental understanding of either physics, computation or logic must incorporate its new insights into his world view.

Notes & Exercises

(1) I always found it an interesting coincidence that the two basic ingredients of modern quantum theory, namely probability and complex numbers, were discovered by the same person, an extraordinary man of many talents, a gambling scholar by the name of Girolamo Cardano (1501–1576).

(2) Complex numbers have many applications in physics, however, not until the advent of quantum theory was their ubiquitous and fundamental role in the description of the actual physical world so evident. Even today, their profound link with probabilities appears to be a rather mysterious connection. Mathematically speaking, the set of complex numbers is a field. This is an important algebraic structure used in almost all branches of mathematics. You do not have to know much about algebraic fields to follow these lectures, but still, you should know the basics. Look them up.

(3) (a) The sets of rational and real numbers are all fields, but the set of integers is not. Why?
(b) What does it mean that the field of complex numbers is algebraically closed?
(c) Evaluate each of the following quantities $1+e^{-i\pi}$, $|1+i|$, $(1+i)^{42}$, $\sqrt{i}$, 2^i and i^i.
(d) Here is a simple proof that $+1=-1$,

$$1=\sqrt{1}=\sqrt{(-1)(-1)}=\sqrt{-1}\sqrt{-1}=i^2=-1$$

What is wrong with it?

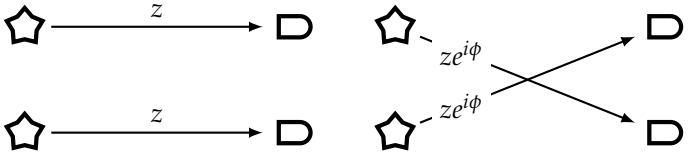
(4) A quantum computer starts calculations in some initial state, then follows n different computational paths which lead to the final output. The computational paths are followed with probability amplitudes $\frac{1}{\sqrt{n}}e^{ik\varphi}$, where φ is a fixed angle $0<\varphi<2\pi$ and $k=0,1,...n-1$. Show that the probability of generating the output is

$$\frac{1}{n}\left|\frac{1-e^{in\varphi}}{1-e^{i\varphi}}\right|^2=\frac{1}{n}\frac{\sin^2(n\frac{\varphi}{2})}{\sin^2(\frac{\varphi}{2})}.$$

for $0<\varphi<2\pi$ and 1 for $\varphi=0$. Plot the probability as a function of φ.

$1+z+z^2+\ldots+z^n=\frac{1-z^{n+1}}{1-z}$

(5) Imagine two distant stars that emit *identical* photons. If you point a single detector towards them you will register a click every now and then, but you never know which star the photon came from. Now prepare two detectors and point them towards the stars. Assume the photons arrive with the probability amplitudes specified below.

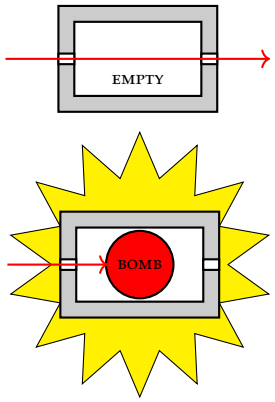


Every now and then you will register a coincidence – the two detectors will fire.
(a) Calculate the probability of a coincidence.
(b) Now, assume that $z\approx\frac{1}{r}e^{i\frac{2r\pi}{\lambda}}$, where r is the distance between detectors and the stars. How can we use this to measure r?

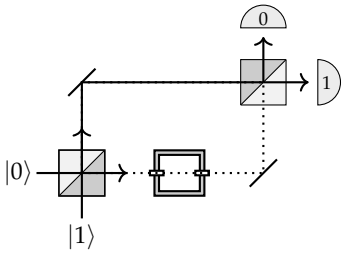
(6) **Quantum Bomb Tester** You have been drafted by the government to help in the demining effort in a former war-zone. In particular, retreating forces have left very sensitive bombs in some of the sealed rooms. The bombs are configured such that if even one photon of light is absorbed by the fuse (i.e. if someone looks into the room), the bomb will go off. Each room has an input and output port which can be hooked up to external devices. An empty room

This is a slightly modified version of a bomb testing problem described by Avshalom Elitzur and Lev Vaidman in *Quantum-mechanical interaction-free measurement*, Found. Phys. **47**, 987-997 (1993).

will let light go from the input to the output ports unaffected, whilst a room with a bomb will explode if light is shone into the input port and the bomb absorbs even just one photon. Your task is to find a way of determining whether a room has a bomb in it without blowing it up, so that specialised (limited and expensive) equipment can be devoted to defusing that particular room. You would like to know with certainty whether a particular room had a bomb in it.



Hint: Consider the setup where the input and output ports are hooked up in one of the arms of a Mach-Zehnder interferometer.



(a) To start with, consider the setup (see the margin) where the input and output ports are hooked up in the lower arm of a Mach-Zehnder interfereometer.
   (i) Assume an empty room. Send a photon to input port $|0\rangle$. Which detector, at the output port, will register the photon?
   (ii) Now assume that the room does contain a bomb. Again, send a photon to input port $|0\rangle$. Which detector will register the photon and with which probability?
   (iii) Design a scheme that allows you – at least part of the time – to decide whether a room has a bomb in it without blowing it up. If you iterate the procedure, what is its overall success rate for the detection of a bomb without blowing it up?

(b) Assume that the two beam splitters in the interferometer are different. Say the first beamsplitter reflects incoming light with probability r and transmits with probability $t = 1 - r$ and the second one transmits with probability r and reflects with probability t. Would the new setup improve the overall success rate of the detection of a bomb without blowing it up?

(c) There exists a scheme, involving many beamsplitters and something called "quantum Zeno effect", such that the success rate for detecting a bomb without blowing it up approaches 100%. Try to work it out or find a solution on internet.

(7) A quantum machine has N perfectly distinguishable configurations. What is the maximum number of computational paths connecting a specific input with a specific output after k steps of the machine? Suppose you are using your laptop to add together amplitudes pertaining to each of the paths. As k and N increase you may need more time and more memory to complete the task. How does the execution time and the memory requirements grow with k and N? Will you need more time or more memory or both?

(8) The classical theory of computation is essentially the theory of the universal Turing machine - the most popular mathematical model of classical computation. Its significance relies on the fact that given a large but finite amount of time the universal Turing machine is capable of any computation that can be done by any modern classical digital computer, no matter how powerful. The concept of Turing machines may be modified to incorporate quantum computation, but we will not follow this path. It is much easier to explain the essence of quantum computation talking about quantum logic gates and quantum Boolean networks or circuits. The two approached are computationally equivalent, even though certain theoretical concepts, e.g. in computational complexity, are easier to formulate precisely using the Turing machine model. The main advantage of quantum circuits is that they relate far more directly to proposed experimental realisations of quantum computation.

(9) In computational complexity the basic distinction is between polynomial versus exponential algorithms. Polynomial growth is good and exponential growth is bad, especially if you have to pay for it. There is an old story about the legendary inventor of chess who asked the Persian king to be paid only by a grain of cereal, doubled on each of the 64 squares of a chess board. The king placed one grain of rice on the first square, two on the second, four on the third, and he was supposed to keep on doubling until the board was full. The

One light year (the distance that light travels through a vacuum in one year) is 9.4607×10^{15}m.

last square would then have $2^{63} = 9,223,372,036,854,775,808$ grains of rice, more than has been ever harvested on planet Earth, to which we must add the grains of all previous squares, making the total number about twice as large. If we placed that many grains in an unbroken line we would reach the nearest star Alpha Centauri, our closest celestial neighbour beyond the solar system, about 4.4 light-years away. The moral of the story: if whatever you do requires an exponential use of resources you are in trouble.

(10) In order to make qualitative distinctions between how different functions grow we will often use the asymptotic big-O notation. For example, suppose an algorithm running on input of size n takes $an^2 + bn + c$ elementary steps, for some positive constants a, b and c. These constants depend mainly on the details of the implementation and the choice of elementary steps. What we really care about is that for large n the whole expression is dominated by its quadratic term. We then say that the running time of this algorithm grows as n^2, and we write it as $O(n^2)$, ignoring the less significant terms and the constant coefficients. More precisely, let $f(n)$ and $g(n)$ be functions from positive integers to positive reals. You may think of $f(n)$ and $g(n)$ as the running times of two algorithms on inputs of size n. We say $f = O(g)$, which means that f grows no faster than g, if there is a constant $c > 0$ such that $f(n) \leq cg(n)$ for all sufficiently large values of n. Essentially, $f = O(g)$ is a very loose analog of $f \leq g$. In addition to the big-O notation, computer scientists often use Ω for lower bounds: $f = \Omega(g)$ means $g = O(f)$. Again, this is a very loose analog of $f \geq g$.

$f = O(g)$ is pronounced as "f is big-oh of g".

(11) (a) When we say that $f(n) = O(\log n)$, why don't we have to specify the base of the logarithm?
(b) Let $f(n) = 5n^3 + 1000n + 50$, is $f(n) = O(n^3)$ or $O(n^4)$ or both?
(c) Which of the following statements are true?
(i) $n^k = O(2^n)$ for any constant k
(ii) $n! = O(n^n)$
(iii) if $f_1 = O(g)$ and $f_2 = O(g)$ then $f_1 + f_2 = O(g)$

(12) There exists a randomised algorithm which tests whether a given number N is prime. The algorithm always returns YES when N is prime and the probability it returns YES when N is not prime is ϵ, which not greater than half (independently, each time you run the algorithm). You run this algorithm (for the same N) r times and each time the algorithm returns YES. What is the probability that N is not prime?

Primality used to be given as the classic example of a problem in BPP but not P. However, in 2002 a deterministic polynomial time test for primality was proposed by Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. Thus, since 2002, primality has been in P.

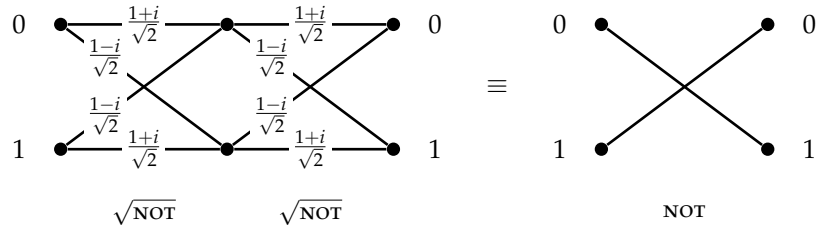
(13) Suppose a randomised algorithm solves a decision problem, returning YES or NO answers. It gets the answer wrong with a probability not greater than $\frac{1}{2} - \delta$, where $\delta > 0$ is a constant.
(a) If we perform this computation r times, how many possible sequences of outcomes are there?
(b) Give a bound on the probability of any particular sequence with w wrong answers.
(c) If we look at the set of r outcomes, we will determine the final outcome by performing a majority vote. This can only go wrong if $w > r/2$. Give an upper bound on the probability of any single sequence that would lead us to the wrong conclusion.
(d) Using the bound $1 - x \leq e^{-x}$, conclude that the probability of our coming to the wrong conclusion is upper bounded by $e^{-2r\delta^2}$.

Chernoff Bound

This result is known as the Chernoff bound. If we are willing to accept a probability of error no larger than ϵ, then it suffices to run the computation a number of times $r = O(\log 1/\epsilon)$.

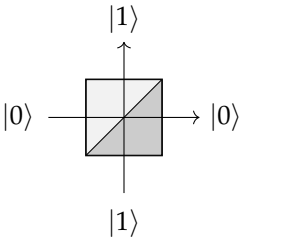
Appendix: Physics against logic
Explained with a beamsplitter

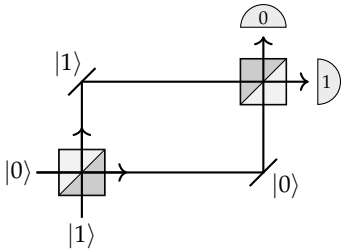
Consider the following task: design a logic gate that operates on a single bit such that when it is followed by another, identical, logic gate the output is always the negation of the input. Let us call this logic gate the square root of NOT ($\sqrt{\text{NOT}}$). A simple check, such as an attempt to construct a truth table, should persuade you that there is no such operation in logic. It may seem reasonable to argue that since there is no such operation in logic, $\sqrt{\text{NOT}}$ is impossible. But think again.

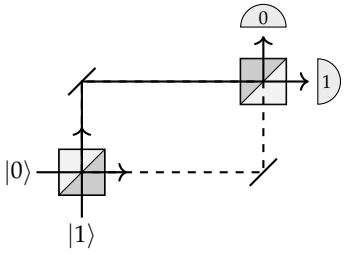


Here is a simple computation, two identical computational steps performed on two states labelled as 0 and 1, i.e. on one bit. An interplay of constructive and destructive interference makes some transitions impossible and the result is the logical NOT. Thus, quantum theory declares, the square root of NOT is possible. And it does exist! Experimental physicists routinely construct this and many other "impossible" gates in their laboratories. In fact, the square root of NOT can be as simple as a symmetric beam-splitter.

A symmetric beam-splitter is a cube of glass which reflects half the light that impinges upon it, while allowing the remaining half to pass through unaffected. For our purposes it can be viewed as a device which has two input and two output ports which we label as $|0\rangle$ and $|1\rangle$. When we aim a single photon at such a beam-splitter using one of the input ports, we notice that the photon doesn't split in two: we can place photo-detectors wherever we like in the apparatus, fire in a photon, and verify that if any of the photo-detectors registers a hit, none of the others do. In particular, if we place a photo-detector behind the beam-splitter in each of the two possible exit beams, the photon is detected with equal probability at either detector, no matter whether the photon was initially fired from input port $|0\rangle$ or $|1\rangle$. It may seem obvious that at the very least, the photon is *either* in the transmitted beam $|0\rangle$ *or* in the reflected beam $|1\rangle$ during any one run of this experiment. Thus we may be tempted to think of the beam-splitter as a random binary switch which, with equal probability, transforms any binary input into one of the two possible outputs. However, that is not necessarily the case. Let us introduce a second beam-splitter and place two normal mirrors so that both paths intersect at the second beam-splitter (see diagrams in the margin).

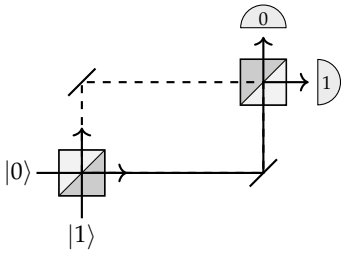






Two consecutive reflections give amplitude $\frac{i}{\sqrt{2}}\frac{i}{\sqrt{2}} = -\frac{1}{2}$

Now, the axiom of additivity in probability theory, says that whenever something can happen in several alternative ways we add probabilities for each way considered separately. We might argue that a photon fired into the input port $|0\rangle$ can reach the detector 0 in two *mutually exclusive* ways: either by two consecutive reflections or by two consecutive transmissions. Each reflection happens with probability $1/2$ and each transmission happens with probability $1/2$ thus the total probability of reaching detector 0 is a sum of the probability of the two consecutive reflections ($1/2 \times 1/2 = 1/4$) and the probability of the two consecutive transmissions ($1/2 \times 1/2 = 1/4$) which gives probability $1/2$. This makes perfect sense – a random switch followed by a random switch should give nothing else but a random switch. However, if we set up such an experiment, that is not what happens! When the optical paths between the two beam-splitters are the same, the photon fired from input port $|0\rangle$ *always* strikes detector 1 and *never* detector 0 (and the photon fired from input port $|1\rangle$ *always* strikes detector 0 and *never* detector 1). Thus a beam-splitter acts as the square root of NOT gate.

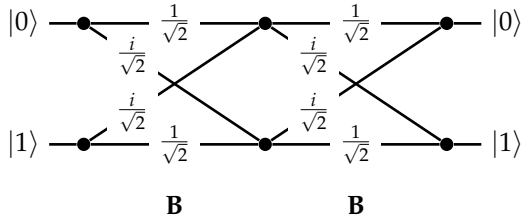


The action of the beamsplitter – in fact, the action of any quantum device – can be described by tabulating the amplitudes of transitions between its input and output ports.

$$B = \begin{bmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

There is no reason why probability theory or any other a priori mathematical construct should make any meaningful statements about outcomes of physical experiments.

The matrix element B_{lk}, where $k, l = 0, 1$, represents the amplitude of transition from input $|k\rangle$ to output $|l\rangle$ (watch the order of indices). Each reflection (entries B_{01} and B_{10}) happens with amplitude $i/\sqrt{2}$ and each transmission (entries B_{00} and B_{11}) happens with amplitude $1/\sqrt{2}$. Thus the total amplitude that a photon fired from input port $|0\rangle$ will reach detector 0 is the sum of the amplitude of the two consecutive reflections ($i/\sqrt{2} \times i/\sqrt{2} = -1/2$) and the amplitude of the two consecutive transmissions ($1/\sqrt{2} \times 1/\sqrt{2} = 1/2$) which gives the total amplitude 0. The resulting probability is then zero. Unlike probabilities, amplitudes can cancel out each other out. We can now go on and calculate the amplitude that the photon will reach detector 1. In this case we will get i, which gives probability 1. We can then switch to input $|1\rangle$ and repeat our calculations. All possible paths and associated amplitudes are shown in the diagram below.



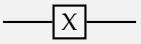
However, instead of going through all the paths in this diagram and linking specific inputs to specific outputs, we can simply multiply the transition matrices,

$$BB = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} = iX.$$

As you can see, the matrix multiplication in one swoop takes care of multiplication and addition of amplitudes corresponding to different alternatives. You can now inform you colleagues logicians that they are now entitled to propose a new logical operation $\sqrt{\text{NOT}}$ for a faithful physical model for it exists in nature!

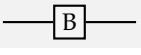
LOGICAL NOT

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$



BEAM SPLITTER

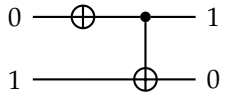
$$B = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$$



Note that gate B is not the same square root of NOT as the one described in the first diagram in this section. There are infinitely many ways of implementing this "impossible" logical operation.

Appendix: Classical world of bits and gates

From a mathematical perspective, computation is an operation on abstract symbols. Any finite set of symbols is called an alphabet and any finite sequence of symbols from that alphabet is called a string. Here, without any loss of generality, we will use the binary alphabet $\{0,1\}$ and we shall denote the set of all 2^n possible binary strings of length n as $\{0,1\}^n$. For example, $\{0,1\}^2$ contains the strings $00, 01, 10$ and 11. Operations on bits are often represented as network (or circuit) diagrams. For example, the circuit



operates on two bits. It should be read from left to right. The horizontal line represents a wire, which inertly carries a bit from one operation to another. Various icons placed on the wires represent elementary logical operations known as logic gates. Here, the first bit (counting from the top) is negated with logical NOT, represented as $\oplus$, and then a controlled-NOT is applied to the first and the second bit. The controlled-NOT flips the second (target) bit if the first (control) bit is 1 and does nothing if the control bit is 0.

It is well known that some elementary operations on bits, for example logical NOT and logical AND, are complete, that is, any Boolean function $\{0,1\}^n \mapsto \{0,1\}$ can be expressed in terms of only these two operations. Some of these essential operations on bits, such as bit addition XOR ($\oplus$) and bit multiplication AND ($\times$), defined as

$$\begin{array}{llll} 0\oplus 0=0 & 0\oplus 1=1 & 1\oplus 0=1 & 1\oplus 1=0, \\ 0\times 0=0 & 0\times 1=0 & 1\times 0=0 & 1\times 1=1, \end{array}$$

are irreversible – given the output, we usually cannot reconstruct the input. However, it is possible to implement these gates, and indeed any function $\{0,1\}^n \mapsto \{0,1\}^m$, using only reversible operations. This is important for us because we will study the physical foundations of computation and the basic laws of physics are reversible in time. This reversibility leads, in a natural way, to reversible logic gates and reversible computation.

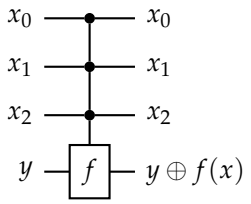
A reversible computer evaluates Boolean functions,

$$f : \{0,1\}^n \mapsto \{0,1\}$$

by embedding them into reversible computation

$$F : (x,y) \mapsto (x, y \oplus f(x)),$$

where $x \in \{0,1\}^n$, $y \in \{0,1\}$. The corresponding circuit shows such a function evaluation for $n = 3$,

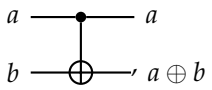


In particular for $y = 0$ we obtain $F(x,0) = (x, f(x))$. This method can easily be generalised to include any function $\{0,1\}^n \mapsto \{0,1\}^m$. We simply embed such a function evaluation into an invertible operation on at least $n+m$ bits. Invertible functions on n bits are effectively permutations of 2^n binary strings of length n. Reversible Boolean function evaluation.

All elementary steps of any reversible computation must be implemented in a reversible way. This is not a problem because all permutations of binary strings can be constructed using a set of simple reversible gates, such as NOT, controlled-NOT (C-NOT) and a controlled-controlled-NOT (C^2-NOT). The controlled-NOT is a reversible

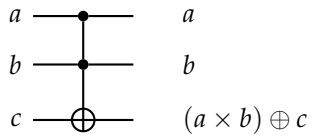
gate which operates on two bits and maps $(a, b) \mapsto (a, a \oplus b)$. It flips the second (target) bit if the first (control) bit is 1 and does nothing if the control bit is 0.

Classical controlled-NOT gate



The c^2-NOT gate, or the Toffoli gate (named after Tommaso Toffoli, a really cool Italian-American computer engineer), is a reversible gate which operates on three bits and maps $(a, b, c) \mapsto (a, b, (a \times b) \oplus c)$. This gate has two control bits a and b and one target bit c, and it flips the target bit if both of the the control bits are 1, and does nothing otherwise.

The Toffoli gate, or the controlled-controlled-NOT



The Toffoli gate is very powerful. If we have a source of 0s and 1s and can choose to ignore some inputs or outputs then we can obtain: negation of c (set $a = b = 1$, a C-NOT gate with b as the control and c as the target (set $a = 1$) or just $b \oplus c$ if we ignore b at the output, and finally an AND gate with inputs a and b (set $c = 0$ and ignore a and b at the output). The Toffoli gate gives us all we need to construct evaluation of any Boolean function.

Reversible computation predates quantum computation. Back in the 1980s Rolf Landauer of IBM, while thinking about the ultimate limits to computation, realised that to erase information costs energy. But if a computer were to operate reversibly, Landauer argued, then in principle there would be no energy dissipation associated with irreversible logic elements and hence and no power requirement to run computation. We can compute for free!

A QUBIT

ARTUR EKERT

Lecture 1

Introduction to Quantum Information Science

About quantum bits, quantum circuits and impossible logic operations, such as $\sqrt{\text{NOT}}$, which nonetheless can be implemented. About single-qubit unitaries and rotations of the Bloch sphere.

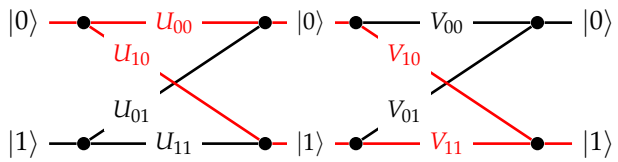
In order to understand something in its full complexity it is always good to start with the simplest case. Let us take a closer look at quantum interference in the simplest possible computing machine, the one that has only two distinguishable configurations — two quantum states — which we label as $|0\rangle$ and $|1\rangle$. We prepare the machine in some input state, usually $|0\rangle$, and let it evolve. The machine undergoes a prescribed sequence of computational steps, each of which induces transitions between the two "computational states", $|0\rangle$ and $|1\rangle$. The machine then ends in the output state $|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle$, meaning the two outputs, $|0\rangle$ and $|1\rangle$, are reached with probability amplitudes α_0 and α_1, respectively. In the process of computation each computational step U sends state $|k\rangle$ to state $|l\rangle$, where $k, l = 0, 1$, but only with some *amplitude* U_{lk}. We write this as

$$|k\rangle \to \sum_l U_{lk} |l\rangle \,. \tag{1}$$

(watch the order of indices). Thus any computational step U of this machine can be described by a matrix which tabulates all the transition amplitudes,

$$U = \begin{bmatrix} U_{00} & U_{01} \\ U_{10} & U_{11} \end{bmatrix}.$$

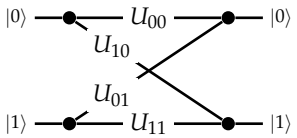
The matrix element U_{lk} represents the amplitude of transition from state $|k\rangle$ to state $|l\rangle$ (again, watch the order of indices). To be sure, the entries in this matrix are not any random complex numbers; their moduli squared represent transition probabilities which in turn implies that such matrices must be unitary (see prerequisite material). Where is the interference? Consider two computational steps, U and V,



U **V**

What is the amplitude that input $|k\rangle$ will generate output $|m\rangle$? We have to check all computational paths leading from input $|k\rangle$ to output $|m\rangle$ and add the corresponding amplitudes. For example, as you can see in the diagram above, input $|0\rangle$ and output $|1\rangle$ are connected by the two computational paths, $|0\rangle \mapsto |0\rangle \mapsto |1\rangle$ (amplitude $V_{10}U_{00}$) and $|0\rangle \mapsto |1\rangle \mapsto |1\rangle$ (amplitude $V_{11}U_{10}$). Thus the total amplitude that input $|0\rangle$ gives ouput $|1\rangle$ is the sum $V_{10}U_{00} + V_{11}U_{10}$, and when we take the mod square of this expression we will see the interference term. In general, given U and V,

$$|k\rangle \to \sum_l U_{lk} |l\rangle \,, \quad |l\rangle \to \sum_m V_{ml} |m\rangle \,, \tag{2}$$



U

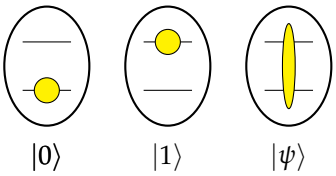
Recall that matrix U is called unitary if $U^\dagger U = UU^\dagger = \mathbb{1}$, where the *adjoint* or *Hermitian conjugate* $U^\dagger$ of any matrix U with complex entries U_{ij} is obtained by taking the complex conjugate of every element in the matrix and then interchanging rows and columns ($U^\dagger_{kl} = U^\star_{lk}$).

we compose the two operations, we apply first U and then V, to obtain

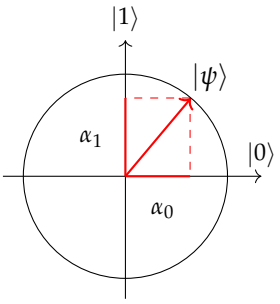
$$|k\rangle \to \sum_l U_{lk} \left(\sum_m V_{ml} |m\rangle \right) = \sum_m \left(\sum_l V_{ml} U_{lk} \right) |m\rangle = \sum_m (VU)_{mk} |m\rangle . \qquad (3)$$

If you want to hone your quantum intuition think about it this way. The amplitude that input $|k\rangle$ evolves to $|m\rangle$ via a specific intermediate state $|l\rangle$ is given by $V_{ml}U_{lk}$ (evolutions are independent so the amplitudes are multiplied). This done we have to sum over all possible values of l (the transition can occur in several mutually exclusive ways so the amplitudes are added) to obtain $\sum_l V_{ml}U_{lk}$. Thus the matrix multiplication VU (watch the order of matrices) in one swoop takes care of multiplication and addition of amplitudes corresponding to different computational paths.

1.1. **Quantum bits called qubits.** A two-state machine that we have just described in abstract terms is usually realised as a controlled evolution of a two state system, called a quantum bit or a qubit. For example, state $|0\rangle$ may be chosen to be the lowest energy state of an atom, the ground state, and state $|1\rangle$ a higher energy state, the excited state. Pulses of light of appropriate frequency, duration and intensity can take the atom back and forth between the basis states $|0\rangle$ and $|1\rangle$ (implementing logical NOT). Some other pulses, say, half the duration or intensity will take the atom into states that have no classical analogue. Such states are called *coherent superpositions* of $|0\rangle$ and $|1\rangle$ and represent a qubit in state $|0\rangle$ with some amplitude α_0 and in state $|1\rangle$ with some other amplitude α_1. This is conveniently represented by a state vector



$$|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle \leftrightarrow \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} . \qquad (4)$$



$|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle$

A *qubit* is a quantum system in which the Boolean states 0 and 1 are represented by a prescribed pair of normalised and mutually orthogonal quantum states labeled as $\{|0\rangle, |1\rangle\}$. The two states form a 'computational basis' or a 'standard basis' and any other state of an isolated qubit can be written as a coherent superposition $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ for some α_0 and α_1 such that $|\alpha_0|^2 + |\alpha_1|^2 = 1$. A qubit is typically a microscopic system, such as an atom, a nuclear spin, or a polarised photon.

As we have already mentioned, any computational step, that is, any physically admissible operation U on a qubit, is described by a 2×2 unitary matrix U. It modifies the state of the qubit

$$|\psi\rangle \to |\psi'\rangle = U |\psi\rangle ,$$

which we can write explicitly as

$$\begin{bmatrix} \alpha'_0 \\ \alpha'_1 \end{bmatrix} = \begin{bmatrix} U_{00} & U_{01} \\ U_{10} & U_{11} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} .$$

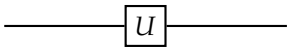
That is, operation U sends state $|\psi\rangle$, with components α_k, into state $|\psi'\rangle = U|\psi\rangle$, with components $\alpha'_l = \sum_k U_{lk}\alpha_k$.

Here we are talking about isolated systems. As you will learn soon, a larger class of physically admissible operations is described by completely positive maps. It may sound awfully complicated but, as you will see soon, it is very simple.

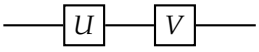
1.2. **Quantum gates and circuits.** Atoms, trapped ions, molecules, nuclear spins and many other quantum objects, which we call qubits, can be used to implement simple quantum interference, and hence simple quantum computation. There is no need to learn about physics behind these diverse technologies if all you want is to understand the basics of quantum computation. We may now conveniently forget about any specific experimental realisation of a qubit and just remember that any manipulations on qubits have to be performed by physically admissible operations, and that such operations are represented by unitary transformations.

A *quantum logic gate* is a device which performs a fixed unitary operation on selected qubits in a fixed period of time and a *quantum circuit* is a device consisting of quantum logic gates whose computational steps are synchronised in time. The *size* of the circuit is the number of gates it contains.

Unitary U acting on a single qubit is represented diagrammatically as

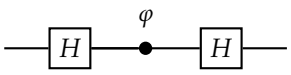


This diagram should be read from left to right. The horizontal line represents a qubit that is inertly carried from one quantum operation to another. We often call this line a quantum wire. The wire may describe translation in space, e.g. atoms traveling through cavities, or translation in time, e.g. a sequence of operations performed on a trapped ion. A sequence of two gates acting on the same qubit, say U followed by V,



is described by the matrix product VU (note the order in which we multiply the matrices).

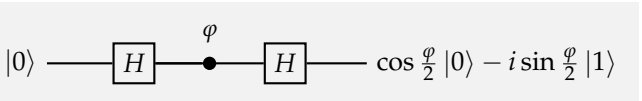
1.3. **Single qubit interference.** Let me now describe what is probably the most important sequence of operations performed on a single qubit, namely a generic single qubit interference. It is typically constructed as a sequence of three elementary operations: the Hadamard gate, followed by a phase shift gate, and followed by the Hadamard gate. We represent it graphically as



You will see it over and over again, for it is quantum interference that gives quantum computation additional capabilities. The product of the three matrices $HP_\varphi H$ describes the action of the whole circuit; it gives the transition amplitudes between states $|0\rangle$ and $|1\rangle$ at the input and the output,

$$e^{i\frac{\varphi}{2}}\begin{bmatrix}\cos\varphi/2 & -i\sin\varphi/2\\ -i\sin\varphi/2 & \cos\varphi/2\end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & e^{i\varphi}\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

Given that our input state is almost always $|0\rangle$ it is sometimes much easier and more instructive to step through the execution of this circuit and follow the evolving state. The interference circuit effects the following sequence of transformations,



$$|0\rangle \overset{H}{\mapsto} \frac{1}{\sqrt{2}}\left(|0\rangle+|1\rangle\right) \overset{P_\phi}{\mapsto} \frac{1}{\sqrt{2}}\left(|0\rangle+e^{i\phi}|1\rangle\right) \overset{H}{\mapsto} \cos\frac{\phi}{2}|0\rangle - i\sin\frac{\phi}{2}|1\rangle\,. \tag{5}$$

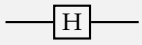
The first Hadamard gate prepares an equally weighted superposition of $|0\rangle$ and $|1\rangle$ and the second one closes the interference by brining the interfering paths together. The phase shift φ effectively controls the evolution and determines the output. The probabilities of finding the qubit in state $|0\rangle$ or $|1\rangle$ at the output are, respectively,

$$\Pr(0)=\cos^2\frac{\phi}{2},\quad \Pr(1)=\sin^2\frac{\phi}{2}. \tag{6}$$

This simple quantum process contains, in a nutshell, the essential ingredients of quantum computation. The sequence, Hadamard - phase shift - Hadamard, will

HADAMARD

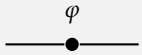
$$H=\tfrac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$



$$|0\rangle \to \tfrac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$$
$$|1\rangle \to \tfrac{1}{\sqrt{2}}(|0\rangle-|1\rangle)$$

PHASE

$$P_\varphi=\begin{bmatrix}1 & 0\\ 0 & e^{i\varphi}\end{bmatrix}$$

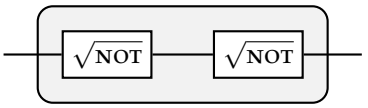


$$|0\rangle \to |0\rangle$$
$$|1\rangle \to e^{i\varphi}|1\rangle$$

We have ignored the global phase factor $e^{i\frac{\varphi}{2}}$

appear over and over again. It reflects a natural progression of quantum computation: first we prepare different computational paths, then we evaluate a function which effectively introduces phase shifts into different computational paths, then we bring the computational paths together at the output.

1.4. **The square root of NOT.** Now that we have poked our heads into the quantum world, let us see how quantum interference challenges conventional logic. Consider a following task: design a logic gate that operates on a single bit and such that when it is followed by another, identical, logic gate the output is always the negation of the input. Let us call this logic gate the square root of NOT ($\sqrt{\text{NOT}}$). A simple check – such as an attempt to construct a truth table – should persuade you that there is no such operation in logic. It may seem reasonable to argue that since there is no such operation in logic, $\sqrt{\text{NOT}}$ is impossible. But it does exist! Experimental physicists routinely construct such "impossible" gates in their laboratories. It is a physically admissible operation described by the unitary



NOT

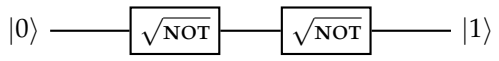
$$\sqrt{\text{NOT}} = \frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\frac{\pi}{4}} & e^{-i\frac{\pi}{4}} \\ e^{-i\frac{\pi}{4}} & e^{i\frac{\pi}{4}} \end{bmatrix}.$$

Indeed,

$$\frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

There are infinitely many unitary operations that effects the square root of NOT.

You can also step through the circuit diagram and follow the evolution of the state vector, e.g.



$$|0\rangle \overset{\sqrt{\text{NOT}}}{\longrightarrow} \frac{1}{\sqrt{2}}\left[e^{i\frac{\pi}{4}}|0\rangle + e^{-i\frac{\pi}{4}}|1\rangle\right] \overset{\sqrt{\text{NOT}}}{\mapsto} |1\rangle\,. \tag{7}$$

If you prefer to work with column vectors and matrices, you can write the two consecutive application of $\sqrt{\text{NOT}}$ to state $|0\rangle$ as

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\frac{\pi}{4}} & e^{-i\frac{\pi}{4}} \\ e^{-i\frac{\pi}{4}} & e^{i\frac{\pi}{4}} \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\frac{\pi}{4}} \\ e^{-i\frac{\pi}{4}} \end{bmatrix} \leftarrow \frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\frac{\pi}{4}} \\ e^{-i\frac{\pi}{4}} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\frac{\pi}{4}} & e^{-i\frac{\pi}{4}} \\ e^{-i\frac{\pi}{4}} & e^{i\frac{\pi}{4}} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Following a well established convention the formulae above should be read from right to left. Confused? Well, you are not the only one. Just remember that circuits diagrams are read from left to right and vector and matrix operations go from right to left. This way or another, quantum theory explains the behaviour of $\sqrt{\text{NOT}}$, hence, reassured by the physical experiments that corroborate this theory, logicians are now entitled to propose a new logical operation $\sqrt{\text{NOT}}$. Why? Because a faithful physical model for it exists in nature!

1.5. **The phase gates galore.** Apart from a generic phase gate P_φ let us mention three specific phase gates that will reoccur frequently

$$P_\varphi = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{bmatrix} \qquad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \qquad T = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{bmatrix}$$

PHASE FLIP PHASE GATE

The phase gate α is defined up to a global phase factor. We can write its matrix as

$$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{bmatrix} \text{ or as } \begin{bmatrix} e^{-i\frac{\varphi}{2}} & 0 \\ 0 & e^{i\frac{\varphi}{2}} \end{bmatrix}$$

The first version is more common in the quantum information science community, but the second one is sometimes more convenient to use for it has determinant 1 and hence belongs to the $SU(2)$ group. We will occasionally switch to the $SU(2)$ version of a phase gates.

The T gate is also known as the $\pi/8$ gate (this is to confuse you completely) for the $SU(2)$ version of the T matrix has $e^{\mp i\pi/8}$ on the diagonal.

1.6. **Pauli gates.** Let us add to our collection of the most common single qubit gates the three Pauli operators (matrices), $\sigma_x \equiv X$, $\sigma_y \equiv Y$, and $\sigma_z \equiv Z$, supplemented by the identity matrix, $\mathbb{1}$,

$$\mathbb{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad Z = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$$

BIT FLIP PHASE FLIP

We use the standard basis $\{|0\rangle, |1\rangle\}$ most of the time and often refer to operators as matrices.

The Pauli matrices are both unitary and Hermitian. They square to the identity and anticommute

$$XY + YX = 0,$$
$$XZ + ZX = 0,$$
$$YZ + ZY = 0,$$

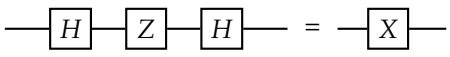
and satisfy

$$XY = iZ$$

(and cyclic permutations)

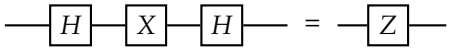
The identity is just a quantum wire and the X gate is the logical NOT, also known as the bit flip. The remaining two gates, Y and Z, do not have classical analogues. The Z gate known as the phase flip, it flips the sign in front of $|1\rangle$, and the Y gate describes the combined effect of both the bit and the phase flip, $ZX = iY$. The Pauli matrices are so ubiquitous in quantum physics, that anyone who wants to work in the field should memorise them. (No, I am not exaggerating.) They are also called sigma matrices or the Pauli spin matrices.

1.7. **From bit flips to phase flips and back again.** The Pauli Z gate is a special case of a phase gate, with $\varphi = \pi$. When we insert it into the interference circuit we obtain,



If you wish to verify this write the Hadamard gate as $H = (X + Z)/\sqrt{2}$ and use the properties of the Pauli operators. The Hadamard gate turn phase flips into bit flips, $HZH = X$, and bit flips into phase flips $HXH = Z$,

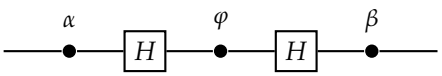
$$HXH = Z$$
$$HZH = X$$
$$HYH = -Y$$



Unitaries, such as H, that take the three Pauli operators to the Pauli operators via conjugation form the Clifford group, which we will meet later on. Which phase gate is in the Clifford group of a single qubit?

Let us also add, for completeness, that $HYH = -Y$. You will see these identities again, and again, when we discuss quantum error corrections.

1.8. **Any unitary operation on a single qubit.** There are infinitely many unitary operations that can be performed on a single qubit. In general, any complex $n \times n$ matrix has n^2 complex entires, hence it can be specified by $2n^2$ real independent parameters. The unitarity constraint removes n^2 of these hence any unitary $n \times n$ matrix has n^2 real independent parameters. In particular, we need four real parameters to specify a 2×2 unitary matrix. If we are prepared to ignore global phase factors, which we are, there are only three real parameters left. Can we construct and implement any unitary on a single qubit in some simple way? Yes, we can. Any unitary operation on a qubit (up to an overall multiplicative phase factor) can be implemented by a circuit containing just two Hadamards and three phase gates, with adjustable phase settings,



(8)

The phase gate α is defined up to a global phase factor. We can write its matrix as

$$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{bmatrix} \text{ or as } \begin{bmatrix} e^{-i\frac{\alpha}{2}} & 0 \\ 0 & e^{i\frac{\alpha}{2}} \end{bmatrix}$$

The first version is more common in the quantum information science community, but the second one is sometimes more convenient to use for it has determinant 1 and hence belongs to the $SU(2)$ group.

If we multiply the matrices corresponding to each gate in the network (remember that the order of matrix multiplication is reversed) we obtain

$$U(\alpha, \beta, \gamma) = \begin{bmatrix} e^{-i\left(\frac{\alpha+\beta}{2}\right)} \cos \varphi/2 & -ie^{i\left(\frac{\alpha-\beta}{2}\right)} \sin \varphi/2 \\ -ie^{-i\left(\frac{\alpha-\beta}{2}\right)} \sin \varphi/2 & e^{i\left(\frac{\alpha+\beta}{2}\right)} \cos \varphi/2 \end{bmatrix}.$$

Any 2×2 unitary matrix (up to global phase) can be expressed in this form using the three independent real parameters, α, β, and φ, which take values from 0 to 2π. In order to see that this construction does what it claims let us explore an

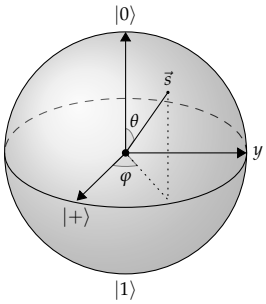
intriguing mathematical connection between single qubit unitaries and rotations in three dimensions.

1.9. Unitary operations on a single qubit form a group. More precisely, the set of all 2×2 unitary matrices forms a non-abeliat group under the matrix multiplication. The group is denoted $U(2)$. It turns out that compositions of single qubit unitaries behave pretty much the same as compositions of rotations in three dimensions. Technically speaking, $U(2)/U(1) \cong SO(3)$, that is, 2×2 unitaries, up to global phase, form a group which is isomorphic to the group of rotations in three dimensions, denoted $SO(3)$. This isomorphism helps to visualise the actions of single qubit gates. There are many ways to introduce this isomorphism. Here we will first show how to represent single-qubit state vectors in terms of Euclidean vectors in three dimensions and then relate unitary operations on state vectors with rotations in the Euclidean space.

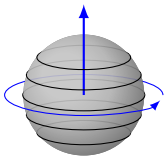
1.10. Bloch sphere. Any single qubit state can be written as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, constrained by the relation $|\alpha|^2 + |\beta|^2 = 1$. This suggests a more natural parametrisation as $|\psi\rangle = \cos\frac{\theta}{2}e^{i\phi_0}|0\rangle + \sin\frac{\theta}{2}e^{i\phi_1}|1\rangle$. (There is a reason we use $\theta/2$ rather than θ, it will be explained later on.) We can factor out a global phase

$$|\psi\rangle = e^{i\phi_0}\left(\cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}e^{i\varphi}|1\rangle\right),$$

and even remove it completely – states that are identical up to a global phase are physically indistinguishable. The parametrisation in terms of θ and φ should remind you of spherical polar coordinates for the surface of a sphere, We call this sphere the *Bloch sphere* and the unit vector $\vec{s}$ defined by θ and φ the Bloch vector. This is a very useful way to visualise quantum states of a single qubit and unitary operations that we perform on it. Any unitary action on the state vector will induce a rotation of the corresponding Bloch vector. But what kind of rotation? Note that any two orthogonal state vectors appear on the Bloch sphere as two Bloch vectors pointing in opposite directions. Now, the two eigenvectors of a single-qubit unitary U must be orthogonal, and so define an axis running through the centre of the Bloch sphere. This is the axis the Bloch vector is rotated about when U acts on the corresponding state vector. The rotation angle α is given by the eigenvalues of U, which, up to a global phase factor, are of the form $e^{\mp i\alpha/2}$.



Phase gates represent rotations about z-axis.

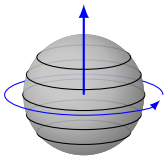


1.11. It is instructive to work out few simple cases and get a feel for the rotations corresponding to the most common unitaries. For example, it is easy to check that a phase gate P_α maps

$$\cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle \mapsto \cos\frac{\theta}{2}|0\rangle + e^{i(\varphi+\alpha)}\sin\frac{\theta}{2}|1\rangle.$$

The azimuthal angle changes from φ to $\varphi + \alpha$, thus the Bloch sphere is rotated anticlockwise by α about the z-axis. The Bloch vectors corresponding to the two eigenvectors of P_α, namely $|0\rangle$ and $|1\rangle$, define the axis of the rotation. Note that the Pauli operator $Z = \sigma_z$ is a special case of a phase gate and represents rotation by 180°, that is π, about z-axis. You can also verify that $X = \sigma_x$, with eigenvectors $(|0\rangle \pm |1\rangle)/\sqrt{2}$, represents rotation by 180° about x-axis, and $Y = \sigma_y$, with eigenvectors $(|0\rangle \pm i|1\rangle)/\sqrt{2}$, represents rotation by 180° about y-axis. How about the Hadamard gate? Like the Pauli operators it squares to the identity $H^2 = \mathbb{1}$, which implies that its eigenvalues are ± 1. Thus it will icorrespond to a rotation by 180°. But about which axis? This time, rather than finding eigenvectors of H, we notice that $HXH = Z$ and $HZH = X$ thus H must swap x and z axes, turning rotations about the z-axis into rotations about the x-axis and vice versa. The Hadamard gate must then represent rotation by 180° about the diagonal $x + z$ axis. You may also notice that after this rotation axis y points in the opposite direction, which seems to be related to another identity, $HYH = -Y$. This is not a coincidence. One can

Phase gates represent rotations about z-axis.

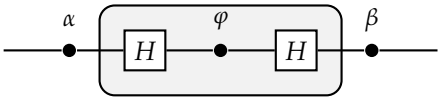


show (exercises) that the effect of the rotation represented by unitary U on the Bloch vector with components s_x, s_y, s_z is summarised in the formula

$$U(s_x X + s_y Y + s_z Z)U^\dagger = s'_x X + s'_y Y + s'_z Z,$$

where s_x, s_y, s_z are the components of the rotated Bloch vector.

1.12. Composition of rotations. We are now in a position understand the circuit (8) in geometric terms. Recall that *any* rotation in the Euclidean space can be performed as a sequence of three rotations: one about z axis, one about x axis and again one about z axis. The circuit does exactly this



The first phase gate effects rotation by α about the z-axis, the second phase gate is sandwiched between the two Hadamard gates and the three gates effect rotation by φ about the x-axis, and finally, the third phase gates effects rotation by β about the z-axis. Thus we can implement any unitary U by choosing the three phase shifts, α, φ and β, known otherwise as the three Euler's angles.

1.13. Finite set of universal gates. The Hadamard and phase gates, with adjustable phases, allow us to implement an arbitrary single-qubit unitary *exactly* . The tacit assumption here is that we have here infinitely many phase gates, one gate for each phase. In fact, we can pick just one phase gate, namely any phase gate P_α with the phase α that is incommensurate with π. It is clear that repeated iteration of P_α can be used to approximate any other phase gate to arbitrary accuracy. Indeed, rotate the Bloch sphere by α about the z-axis sufficiently many times and you end up as close as you please to any other rotation about the z-axis. If you want to be ϵ-close to the desired angle of rotation you may need to repeat rotation by α roughly $1/\epsilon$ times. Indeed, within n applications ($n\alpha \gg 2\pi$) of P_α, we expect the accessible angles to be approximately evenly distributed within the range 0 to 2π, i.e. any angle of rotation can be achieved to an accuracy of $\epsilon = 2\pi/n$ by using up to $n \approx 1/\epsilon$ applications of P_α. Thus we can use just one phase gate to approximate the three phase gates in the circuit (8). There are other ways of implementing irrational rotations of the Bloch sphere. For example, take the Hadamard and the T gate. You can check that the compositions $THTH$ and $HTHT$ represent rotations by angles which are irrational multiples of π, about two different axes. You can then compose a sequence of these two rotations to approximate any other rotation of the sphere (see exercises). This may look very nice in theory but there are issues with this approach. All the gates in the circuit will operate with final precision and the phase gates will deviate from implementing the required irrational rotations. It turns out, however, that we can tolerate minor imperfections; the final result will not be that far off (exercises).

Self-Guided Tour: Ubiquitous Pauli matrices, rotations and single qubit unitaries

Let me elaborate here on relations among single qubit unitary transformations, Pauli matrices and rotations of regular three dimensional vectors. I want you to be able to visualise sequences of unitary operations on a qubit as sequences of rotations, and to see the action of some quantum circuits without getting engaged in lengthy calculations. Matrices form a vector space for you can add them and you can multiply them by a scalar. One possible choice of a basis in this vector space is a set of matrices with a single entry equal to 1 and all other entries 0. However, it turns out that for 2×2 matrices there is another basis — the three Pauli matrices and the identity — which offers lots of insights into the structure of the general single qubit unitary transformations.

$$\mathbb{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

The three Pauli matrices $\sigma_1 \equiv \sigma_x \equiv X$, $\sigma_2 \equiv \sigma_y \equiv Y$, and $\sigma_3 \equiv \sigma_z \equiv Z$, here supplemented by the identity matrix $\mathbb{1}$ (sometimes denoted by σ_0), are both unitary and Hermitian. They square to the identity and anticommute. Any 2×2 complex matrix A has a unique expansion of the form,

The Pauli matrices square to the identity and anticommute
$$XY + YX = 0,$$
$$XZ + ZX = 0,$$
$$YZ + ZY = 0,$$
and satisfy
$$XY = iZ$$
(and cyclic permutations)

$$A = \begin{bmatrix} a_0 + a_z & a_x - ia_y \\ a_x + ia_y & a_0 - a_z \end{bmatrix} = a_0\mathbb{1} + a_x X + a_y Y + a_z Z = a_0\mathbb{1} + \vec{a}\cdot\vec{\sigma}. \tag{9}$$

for some complex numbers a_0, a_x, a_y and a_z. Here $\vec{a}$ is a vector with three complex components (a_x, a_y, a_z) and $\vec{\sigma}$ represents the "vector" of Pauli matrices $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. The algebraic properties of the Pauli matrices can be neatly compacted into a single expression, the multiplication rule (see exercises),

$$(\vec{a}\cdot\vec{\sigma})(\vec{b}\cdot\vec{\sigma}) = (\vec{a}\cdot\vec{b})\ \mathbb{1} + i(\vec{a}\times\vec{b})\cdot\vec{\sigma}. \tag{10}$$

We now introduce the inner product of two matrices, A and B, as $(A|B) = \frac{1}{2}\mathrm{Tr}\, A^\dagger B$ known as the *Hilbert-Schmidt product*.

The trace of a square matrix A, denoted by $\mathrm{Tr}\, A$, is defined to be the sum of the elements on the main diagonal of A. The trace is a linear mapping: for any scalars α and β
$\mathrm{Tr}\,(\alpha A + \beta B) = \alpha \mathrm{Tr}\, A + \beta \mathrm{Tr}\, B$.
Moreover,
$\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)$
$\mathrm{Tr}\,(ABC) = \mathrm{Tr}\,(CAB) = \mathrm{Tr}\,(BCA)$
(all cyclic permutations)

(1) Show that the identity and the three Pauli matrices form an orthonormal basis with respect to the Hilbert-Schmidt product in the space of complex 2×2 matrices.
(2) Show that the coefficients a_k in Eq. (9) are given by the inner products $a_k = (\sigma_k|A) = \frac{1}{2}\mathrm{Tr}\,\sigma_k A$.
(3) Show that $\frac{1}{2}\mathrm{Tr}(\vec{a}\cdot\vec{\sigma})(\vec{b}\cdot\vec{\sigma}) = \vec{a}\cdot\vec{b}$.
(4) Show that for any $\vec{n}\cdot\vec{\sigma}$, the eigenvalues are $\pm|\vec{n}|$.
(5) Show that if $\vec{n}\cdot\vec{m} = 0$, the operators $\vec{n}\cdot\vec{\sigma}$ and $\vec{m}\cdot\vec{\sigma}$ anticommute.

Here we will usually deal with matrices which are Hermitian ($A = A^\dagger$) or unitary ($AA^\dagger = \mathbb{1}$). It is easy to see that if A is Hermitian then both a_0 and the three components of $\vec{a}$ are real. The 2×2 unitaries are usually parametrised as

$$U = e^{i\gamma}(u_0\mathbb{1} + i(u_x X + u_y Y + u_z Z)), \tag{11}$$

where $e^{i\gamma}$ is an overall multiplicative phase factor, with real γ, and both u_0 and the three components u_x, u_y, u_z are real numbers such that $u_0^2 + u_x^2 + u_y^2 + u_z^2 = 1$.

(6) Show that the unitarity condition implies $u_0^2 + u_x^2 + u_y^2 + u_z^2 = 1$.
(7) Show that the determinant of U is $e^{i2\gamma}$.

Geometrically speaking the group of unitaries $U(2)$ is a three-dimensional sphere S^3 in $\mathbb{R}^4$. We often fix the determinant to be $+1$ and express 2×2 unitaries as

$$U = u_0 \mathbb{1} + i(u_x X + u_y Y + u_z Z). \tag{12}$$

Such matrices form a popular subgroup of $U(2)$; it is called the special (meaning determinant is 1) unitary group, $SU(2)$. In quantum theory any two unitary matrices that differ by some global multiplicative phase factor represent the same physical operation, thus we could fix the determinant and restrict ourselves to the $SU(2)$ matrices. This is a sensible approach, practiced by many theoretical physicists, but again, for some historical reasons the convention in quantum information science does not follow this approach. For example, phase gates are usually written as

$$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{bmatrix} \qquad \text{rather than} \qquad \begin{bmatrix} e^{-i\frac{\alpha}{2}} & 0 \\ 0 & e^{i\frac{\alpha}{2}} \end{bmatrix}.$$

Still, sometimes the T gate diag $[1, e^{i\pi/4}]$ is called the $\pi/8$ gate because of its $SU(2)$ form diag$[e^{-i\pi/8}, e^{i\pi/8}]$.

Let us write any 2×2 unitary, to within an overall phase factor, as

$$U = u_0 \mathbb{1} + i(u_x X + u_y Y + u_z Z) = u_0 \mathbb{1} + i\vec{u} \cdot \vec{\sigma}$$

where $u_0^2 + |\vec{u}|^2 = 1$. This additional unitarity restriction allows us to parametrise u_0 and $\vec{u}$ in terms of a real unit vector $\vec{n}$, parallel to $\vec{u}$, and a real angle θ so that

As you can see, we often make progress and gain insights by choosing a convenient parametrisation.

$$U = \cos\theta \; \mathbb{1} + i \sin\theta \; \vec{n} \cdot \vec{\sigma}$$

An alternative way of writing this expression is

$$U = e^{i\theta \vec{n} \cdot \vec{\sigma}}.$$

This follows from the power-series expansion of the exponential. Indeed, any unitary matrix can always be written in the exponential form as,

$$e^{iA} \equiv \mathbb{1} + iA + \frac{(iA)^2}{1 \cdot 2} + \frac{(iA)^3}{1 \cdot 2 \cdot 3} \ldots = \sum_{n=0}^{\infty} \frac{(iA)^n}{n!}, \tag{13}$$

where A is a Hermitian matrix. This is analogous to writing complex numbers of unit moduli in the polar form as $e^{i\alpha}$.

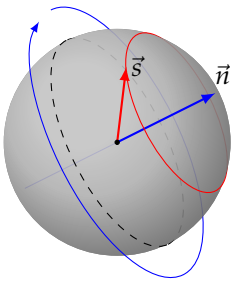
(8) Show that if A squares to the identity, $A^2 = \mathbb{1}$, we can turn the power series expansion into a simple expression; for any real α,

$$e^{i\alpha A} = \cos\alpha \; \mathbb{1} + i \sin\alpha \; A.$$

(9) Using this result, or otherwise, show that any 2×2 unitary matrix U can be written, up to an overall multiplicative phase factor, as

$$U = e^{i\theta \vec{n} \cdot \vec{\sigma}} = \cos\theta \; \mathbb{1} + i \sin\theta \; \vec{n} \cdot \vec{\sigma},$$

The argument here is the same as the argument that $e^{i\theta} = \cos\theta + i \sin\theta$.



$e^{i\theta \vec{n} \cdot \vec{\sigma}}$ rotates vector $\vec{s}$ about $\vec{n}$ by angle 2θ

Here comes a remarkable connection between two-dimensional unitary matrices and ordinary three-dimensional rotations. The unitary $U = e^{i\theta \vec{n} \cdot \vec{\sigma}}$ represents a clockwise rotation about the axis $\vec{n}$ through the angle 2θ (yes, it is two times θ). For example,

$$e^{i\theta\sigma_x} = \begin{bmatrix} \cos\theta & i\sin\theta \\ i\sin\theta & \cos\theta \end{bmatrix} \quad e^{i\theta\sigma_y} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}, \quad e^{i\theta\sigma_z} = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$$

represent rotations by 2θ about the x, y and z-axis, respectively. The Hadamard gate

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{\sqrt{2}}(\sigma_x + \sigma_z) = (-i) e^{i\frac{\pi}{2}\frac{1}{\sqrt{2}}(\sigma_x + \sigma_z)},$$

which, up to an overall multiplicative phase factor $(-i)$, is equal to $e^{i\frac{\pi}{2}\frac{1}{\sqrt{2}}(\sigma_x+\sigma_z)}$, represents rotation about the diagonal $x+z$ axis through the angle π.

In somewhat abstract terms, we make the connection between unitaries and rotations by looking how the unitary group $U(2)$ acts on three dimensional Euclidian space of 2×2 Hermitian matrices with *zero trace*. All such matrices S can be written as $S=\vec{s}\cdot\vec{\sigma}$ for some real $\vec{s}$, i.e. each matrix is represented by a Euclidean vector $\vec{s}$ in $\mathbb{R}^3$. Now, $U\in U(2)$ acts on the Euclidean space of matrices S by the rule,

$$S\mapsto S'=USU^\dagger \qquad \text{that is} \qquad \vec{s}\cdot\vec{\sigma}\mapsto\vec{s}'\cdot\vec{\sigma}=U(\vec{s}\cdot\vec{\sigma})U^\dagger, \tag{14}$$

mapping vector $\vec{s}$ to vector $\vec{s}'$. This is a linear map $\mathbb{R}^3\mapsto\mathbb{R}^3$ given by some real 3×3 matrix R. We notice that this map is an isometry (a distance preserving operation) for it preserves the scalar product in the Euclidean space; for any two vectors $\vec{s}$ and $\vec{v}$,

$$\vec{s}'\cdot\vec{v}'=\tfrac{1}{2}\mathrm{Tr}\,[S'V']=\tfrac{1}{2}\mathrm{Tr}\,[(USU^\dagger)(UVU^\dagger)]=\tfrac{1}{2}\mathrm{Tr}\,[SV]=\vec{s}\cdot\vec{v},$$

where $S=\vec{s}\cdot\vec{\sigma}$, $V=\vec{v}\cdot\vec{\sigma}$ and we have used the cyclic property of the trace. This means that matrix R is orthogonal. Furthermore, we can show that $\det R=1$. The only isometries in three dimensional Euclidian space, which are described by orthogonal matrices R with $\det R=1$ are rotations. Thus, in the mathematical lingo, we have established a homomorphism between $U(2)$ and $SO(3)$, where $SO(3)$ stands for the special orthogonal group in three dimensions (the group of all rotations about the origin of three-dimensional Euclidean space $\mathbb{R}^3$ under the operation of composition). It is quite clear from the formula (14) that unitary matrices that differ only by a global multiplicative phase factor, e.g. U and $e^{i\gamma}U$, represent the same rotation.

Orthogonal transformations preserve the length of vectors as well as the angles between them.

Recall that a homomorphism is a structure-preserving map between two algebraic structures of the same type, in our case two groups. An isomorphism between algebraic structures of the same type is one-to-one homomorphism.

Physicists prefer a more direct demonstration, which goes roughly like this. Consider the map $\vec{s}\mapsto\vec{s}'$ induced by $U=e^{i\alpha\vec{n}\cdot\vec{\sigma}}$. For small values of α we can write

$$\vec{s}'\cdot\vec{\sigma}=U(\vec{s}\cdot\vec{\sigma})U^\dagger=(\mathbb{1}+i\alpha(\vec{n}\cdot\vec{\sigma})+\ldots)(\vec{s}\cdot\vec{\sigma})(\mathbb{1}-i\alpha\,(\vec{n}\cdot\vec{\sigma})+\ldots).$$

To the first order in α the formula reads

$$\vec{s}'\cdot\vec{\sigma}=(\vec{s}+2\alpha\,(\vec{n}\times\vec{s}))\cdot\vec{\sigma} \qquad \text{i.e.} \qquad \vec{s}'=\vec{s}+2\alpha\,(\vec{n}\times\vec{s}),$$

which we recognise as a good old textbook formula for an infinitesimal clockwise rotation of $\vec{s}$ about the axis $\vec{n}$ through the angle 2α.

One can also infer that $\det R=1$ from the fact that any matrix in $U(2)$ can be smoothly connected to the identity.

(10) Show that $\mathrm{Tr}\,\sigma_x\sigma_y\sigma_z=2i$.

(11) Let $U(\vec{e}_k\cdot\sigma_k)U^\dagger=U\sigma_kU^\dagger=\vec{f}_k\cdot\vec{\sigma}$. Here U maps the unit vectors $\vec{e}_x$, $\vec{e}_y$ and $\vec{z}_z$, along the axes x, y and z respectively, into new unit vectors $\vec{f}_x$, $\vec{f}_y$ and $\vec{f}_z$. We already know that in the Euclidean space this transformation is described by a 3×3 orthogonal matrix R. How are the three vectors $\vec{f}_x$, $\vec{f}_y$ and $\vec{f}_z$ related to the entries in matrix R.

(12) Show that $\mathrm{Tr}\,\sigma_x\sigma_y\sigma_z=\mathrm{Tr}\,(\vec{f}_x\cdot\vec{\sigma})(\vec{f}_y\cdot\vec{\sigma})(\vec{f}_z\cdot\vec{\sigma})=2i\det R$, which implies $\det R=1$.

(13) Make use of the orthonormality of the Pauli basis and (14) and show that the elements of the matrix R can be expressed in terms of those of the matrix U, in the form,

$$R_{ij}=\tfrac{1}{2}\mathrm{Tr}\left(\sigma_iU\sigma_jU^\dagger\right). \tag{15}$$

Here i,j take values $1,2,3$ and $\sigma_1\equiv\sigma_x,\sigma_2\equiv\sigma_y,\sigma_3\equiv\sigma_z$.

(14) Show that the phase gate $P_\varphi=\mathrm{diag}[1,e^{i\varphi}]$ represents an anticlockwise rotation about the z-axis through the angle φ. Start with the $SU(2)$ version of the phase gate

$$P_\varphi=e^{-i\frac{\varphi}{2}\sigma_z}=\begin{bmatrix} e^{-i\frac{\varphi}{2}} & 0\\ 0 & e^{i\frac{\varphi}{2}}\end{bmatrix} \quad\longrightarrow\quad R=\begin{bmatrix}\cos\varphi & -\sin\varphi & 0\\ \sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{bmatrix}$$

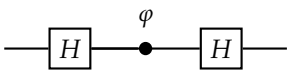
(15) Express the Hadamard, H, in terms of $\vec{n}\cdot\vec{\sigma}$ and demonstrate that

$$HZH = X \qquad HXH = Z \qquad HYH = -Y.$$

(16) Show that the Hadamard gate H turns rotations about the x-axis into rotations about the z-axis and vive versa,

$$H\left(e^{-i\frac{\varphi}{2}Z}\right)H = e^{-i\frac{\varphi}{2}X} \qquad \text{and} \qquad H\left(e^{-i\frac{\varphi}{2}X}\right)H = e^{-i\frac{\varphi}{2}Z}.$$

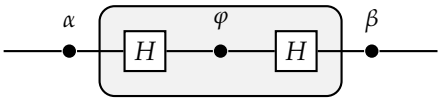
Although this may all seem tediously abstract, it is surprisingly useful. Take another look at the single qubit interference circuit



and the corresponding sequence of unitary operations

$$H\left(e^{-i\frac{\varphi}{2}Z}\right)H = e^{-i\frac{\varphi}{2}X} = \begin{bmatrix} \cos\varphi/2 & -i\sin\varphi/2 \\ -i\sin\varphi/2 & \cos\varphi/2 \end{bmatrix}.$$

The single qubit interference circuit has a simple geometrical meaning. It shows how a rotation about the z-axis, induced by the phase gate P_φ, is turned, by the two Hadamard gates, into a rotation about the x-axis. Now, take a look at this circuit.

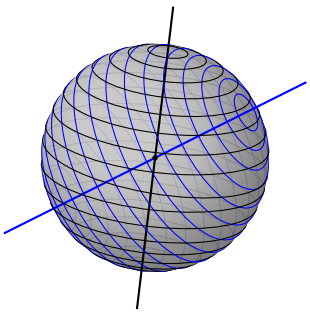


What does it represent? The central part is a rotation by φ about the x-axis, but is it is sandwiched between two rotations about the z-axis. Now I have to appeal to your knowledge of classical mechanics. You may recall that any rotation in the Euclidean space can be performed as a sequence of three rotations: one about z axis, one about x axis and again one about z axis. In this context this implies that any unitary U, up to a global phase factor, can be written as

$$U(\alpha,\beta,\varphi) = e^{-i\frac{\beta}{2}Z}e^{-i\frac{\varphi}{2}X}e^{-i\frac{\alpha}{2}Z} = \begin{bmatrix} e^{-i\left(\frac{\alpha+\beta}{2}\right)}\cos\varphi/2 & ie^{i\left(\frac{\alpha-\beta}{2}\right)}\sin\varphi/2 \\ ie^{-i\left(\frac{\alpha-\beta}{2}\right)}\sin\varphi/2 & e^{i\left(\frac{\alpha+\beta}{2}\right)}\cos\varphi/2 \end{bmatrix}.$$

Thus once you are given a couple of Hadamard gates and an infinite supply of phase gates, so that you can choose the three phases you need, you can construct an arbitrary unitary operation on a single qubit. Needless to say, the two axes, z and x, do not have any special status, geometrically speaking, if we have rotations about any two orthogonal axes we can create any one-qubit unitary that we want.

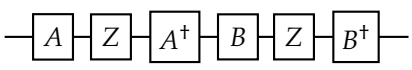
The two axes do not even have to be orthogonal, any two different axes will do. Can you see why?



If we can move along the two families of circles, then from any point on the sphere we can reach any other point.

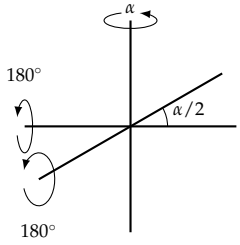
(17) In the Singapore Botanic Gardens, there is a sculpture by Ueli Fausch called "Swiss Granite Fountain". It is a spherical granite ball, which measures 80 cm in diameter and weighs 700 kg, and is kept afloat by strong water pressure directed through the basal block. It is easy to set the ball in motion, and it keeps rotating in whatever way you start for a long time. Suppose you are given access to this ball only near the top, so that you can push it to make it rotate around any horizontal axis, but you don't have enough of a grip to make it turn around the vertical axis. Can you make it rotate around the vertical axis anyway?

Consider the following circuit



Both A and B are unitary operations. We claim that any unitary U can be represented in this form. Again, we can see it geometrically. The circuit represents two rotations by 180° about two axes obtained by rotating the z-axis with unitaries A and B, respectively. Any rotation in the three-dimensional space is the composition

of two rotations by 180°, as shown in the figure. The resulting axis of rotation is perpendicular to the two axes about which rotations by 180° are performed, and the angle of the composed rotation is twice the angle between the two axes.



The time evolution of a quantum state is a unitary process which is generated by a Hermitian operator called the Hamiltonian, H. I hope it will always be clear from the context which H refers to Hamiltonian and which H to Hadamard. Don't confuse the two. The Hamiltonian contains a complete specification of all interactions within the system under consideration. In an isolated system, the state vector $|\psi(t)\rangle$ changes smoothly in time according to the Schrödinger equation

$$\frac{d}{dt}\,|\psi(t)\rangle = -\frac{i}{\hbar}\,H\,|\psi(t)\rangle\,. \tag{16}$$

For time independent Hamiltonians the formal solution reads

$$|\psi(t)\rangle = U(t)\,|\psi(0)\rangle \qquad \text{where} \qquad U(t) = e^{-\frac{i}{\hbar}\,H\,t}. \tag{17}$$

Here $\hbar$ denotes Planck's constant, which has the value $\hbar = 1.05 \times 10^{-34}$ J s. However, theorists always choose to work with a system of units where $\hbar = 1$.

Now, the Hamiltonian of a qubit can always be written as $H = E_0\mathbb{1} + \omega\,\vec{n}\cdot\vec{\sigma}$ hence

$$U(t) = e^{-i\omega t\vec{n}\cdot\vec{\sigma}} = \cos\omega t\;\mathbb{1} - i\sin\omega t\;\vec{n}\cdot\vec{\sigma}. \tag{18}$$

which is a rotation with angular frequency ω about the axis defined by the unit vector $\vec{n}$.

(18) A qubit (spin one-half particle) initially in state $|0\rangle$ (spin up) is placed in a uniform magnetic field. The interaction between the field and the qubit is described by the Hamiltonian

$$H = \omega\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},$$

where ω is proportional to the strength of the field. What is the state of the qubit after time $t = \pi/4\omega$?

In Earth's magnetic field, which is about 0.5 gauss, the value of ω is of the order of 10^6 cycles per second.

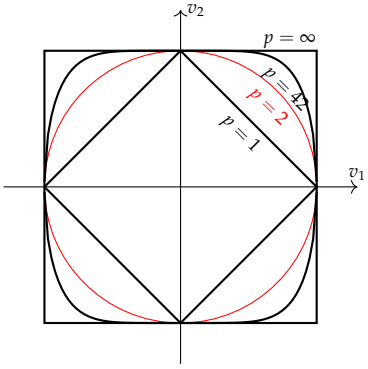
Notes & Exercises

(1) Back in 1926 Max Born simply postulated the connection between amplitudes and probabilities. However, it is worth pointing out, that he did not get it quite right on his first approach. In the original paper proposing the probability interpretation of the state vector (wavefunction) he wrote:

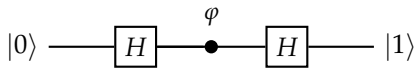
> ...If one translates this result into terms of particles only one interpretation is possible. $\Theta_{\eta,\tau,m}(\alpha,\beta,\gamma)$ [the wavefunction for the particular problem he is considering] gives the probability* for the electron arriving from the z direction to be thrown out into the direction designated by the angles α,β,γ... .
>
> * Addition in proof: More careful considerations show that the probability is proportional to the square of the quantity $\Theta_{\eta,\tau,m}(\alpha,\beta,\gamma)$.

Max Born, Zur Quantenmechanik der Stoßvorgänge, *Zeitschrift für Physik*, 37, 863–867 (1926).

(2) Suppose probabilities are given by the absolute values of amplitudes raised to power p. The admissible physical evolutions must preserve the normalisation of probability. Mathematically speaking, they must be isometries of p-norms. Recall that the p-norm of vector v, with components $v_1, v_2, ..., v_n$, is defined as $\sqrt[p]{|v_1|^p + |v_2|^p + ...|v_n|^p}$. It is clear that any permutation of vector components and multiplication by phase factors (unit complex numbers) will leave any p-norm unchanged. It turns out that these complex permutations are the only isometries, except one special case! For $p = 2$, the isometries are unitary operations, which form a continuous group. In all other cases we are restricted to discrete permutations. We do not have to go into details of the proof for we can *see* this result. The picture in the margin shows unit spheres in different p norms, e.g. for $p = 1, 2, 42$, and ∞. The image of the unit sphere must be preserved under probability preserving operations. As we can see the 2-norm is special because of its rotational invariance – the probability measure picks out no preferred basis in the space of state vectors. Moreover, it respects unitary operations and does not restrict them in any way. If the admissible physical evolutions were restricted to discrete symmetries, e.g. permutations, there would be no continuity and no time as we know it.



(3) **Guess the phase** Consider the usual quantum interference circuit,



Suppose you can control the input of the circuit and measure the output, but you do not know the phase shift φ introduced by the phase gate. You prepare input $|0\rangle$ and register output $|1\rangle$, what can you say about φ? Now you are promised that φ is either 0 or π. You can run the circuit only once to find out which of the two phases was chosen. Can you do that?

(4) Derive the identity

$$(\vec{a}\cdot\vec{\sigma})(\vec{b}\cdot\vec{\sigma}) = (\vec{a}\cdot\vec{b})\,\mathbb{1} + i(\vec{a}\times\vec{b})\cdot\vec{\sigma}.$$

All you need here are the Pauli matrices commutation and anticommutation relations, but it is instructive to derive the identity using the component notation. First, notice that the products of Pauli matrices can be written succinctly as

$$\sigma_i\sigma_j = \delta_{ij}\mathbb{1} + i\varepsilon_{ijk}\,\sigma_k,$$

where δ_{ij} is the Kronecker delta and ε_{ijk} is the Levi-Civita symbol,

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i,j,k) \text{ is } (1,2,3),(2,3,1), \text{ or } (3,1,2), \\ -1 & \text{if } (i,j,k) \text{ is } (3,2,1),(1,3,2), \text{ or } (2,1,3), \\ 0 & \text{if } i=j, \text{ or } j=k, \text{ or } k=i \end{cases}$$

That is, ε_{ijk} is 1 if (i, j, k) is an even permutation of $(1, 2, 3)$, -1 if it is an odd permutation, and 0 if any index is repeated. The Levi-Civita symbol is anti-symmetric, meaning when any two indices are changed, its sign alternates. Then recall that the scalar (dot) product and vector (cross) product of two Euclidean vectors $\vec{a}$ and $\vec{b}$ can be written, in terms of the components, as

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{3} a_i b_i, \qquad (\vec{a} \times \vec{b})_i = \sum_{j,k=1}^{3} \varepsilon_{ijk} a_j b_k.$$

The rest is rather straightforward, $(\vec{a} \cdot \vec{\sigma})(\vec{b} \cdot \vec{\sigma}) = \sum_{ij} a_i b_j \, \sigma_i \sigma_j = \ldots$.

QUANTUM ENTANGLEMENT

ARTUR EKERT

Lecture 2

Introduction to Quantum Information Science

We now know everything we need to know about a single qubit and its quantum behaviour. But, if we want to understand quantum computation, a complicated quantum interference of many interacting qubits, we will need few more mathematical tools. Stepping up from one qubit to two or more is a bigger leap than you might expect. Already with two qubits we will encounter the remarkable phenomenon of quantum entanglement and discuss some of the most puzzling features of quantum theory that took people decades to understand.

The notion of quantum entanglement was the subject of many early debates that focused on the meaning of quantum theory. Back in the 1930s, Albert Einstein, Niels Bohr, Werner Heisenberg and Erwin Schrödinger, to mention only the usual suspects, were trying hard to understand its conceptual consequences. Einstein, the most sceptical of them all, claimed that it was pointing toward the fatal flaw in quantum theory and referred to it as "spooky action-at-a-distance". In contrast, Schrödinger was much more prepared to accept quantum theory exactly as it was formulated and with all its predictions, however weird they might be. In his 1935 paper, which introduced quantum entanglement, he wrote "I would not call it *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought". Today we still talk a lot about quantum entanglement, but more often it is viewed as a physical resource which enables us to communicate with perfect security, build very precise atomic clocks and even teleport small quantum objects! But what exactly is quantum entanglement?

E. Schrödinger, *Discussion of probability relations between separated system*, Mathematical Proceedings of the Cambridge Philosophical Society, Volume 31, Issue 4, October 1935, pp. 555–563. Schrödinger at the time was a fellow of Magdalen College in Oxford.

2.1. **One, two, many...** In classical physics, the transition from a single object to a composite system of many objects is trivial. In order to describe the state of, say, 42 objects at any given moment of time, it is sufficient to describe the state of each of the objects separately. Indeed, the classical state of 42 point-like particles is described by specifying the position and the momentum of each particle. In the classical world the whole is the collection of the parts. The quantum world is quite different. Consider, for example, a pair of qubits. Suppose each of them is described by a state vector, the first one by $|a\rangle$ and the second one by $|b\rangle$. One might therefore think that the most general state of the two qubits should be represented by a pair of state vectors, $|a\rangle\,|b\rangle$, one for each qubit. Such a state is certainly possible but there are other states that cannot be expressed in this form. In order to write down the most general state of two qubits we first focus on the basis states. For a single qubit we have been using the standard basis $\{|0\rangle, |1\rangle\}$. For two qubits we may choose the following as our standard basis states,

$$|0\rangle\,|0\rangle \equiv |00\rangle\,,\ |0\rangle\,|1\rangle \equiv |01\rangle\,,\ |1\rangle\,|0\rangle \equiv |10\rangle\,,\ |1\rangle\,|1\rangle \equiv |11\rangle\,.$$

Within each ket the fist symbol refers to the first qubit and the second to the second and we have tacitly assumed that we can distinguish the two qubits by their location or otherwise. Now, the most general state of the two qubits is a normalised linear combination of these four basis states, that is a vector of the form

$$|\psi\rangle = c_{00}\,|00\rangle + c_{01}\,|01\rangle + c_{10}\,|10\rangle + c_{11}\,|11\rangle\,. \tag{1}$$

Physical interpretation aside, let us count how many real parameters are needed to specify this state. Six, right, for we have four complex numbers (eight real parameters) restricted by the normalisation condition and the fact the states differing only by a global phase factor are equivalent, which leaves us with six real parameters.

Now, by the same line of argument we need only two real parameters to specify a state of a single qubit and hence we need four real parameters to specify any state of two qubits which is of the form $|a\rangle\,|b\rangle$. Thus it cannot be the case that every state of two qubits can be expressed as a pair of states $|a\rangle\,|b\rangle$. For example, compare the two states of two qubits,

$$\frac{1}{\sqrt{2}}\,|00\rangle + \frac{1}{\sqrt{2}}\,|01\rangle \quad \text{and} \quad \frac{1}{\sqrt{2}}\,|00\rangle + \frac{1}{\sqrt{2}}\,|11\rangle\,. \tag{2}$$

The first one is separable for we can view it as a pair of state vectors pertaining to each qubit,

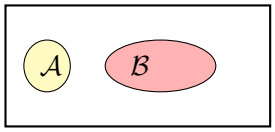
$$\frac{1}{\sqrt{2}}\,|00\rangle + \frac{1}{\sqrt{2}}\,|01\rangle = \underbrace{|0\rangle}_{\text{qubit 1}} \quad \underbrace{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)}_{\text{qubit 2}},$$

whilst the second state does not admit such a decomposition,

$$\frac{1}{\sqrt{2}}\,|00\rangle + \frac{1}{\sqrt{2}}\,|11\rangle \neq |\psi_1\rangle\,|\psi_2\rangle\,, \tag{3}$$

hence we say that it is an entangled state. Informally, any bipartite state that cannot be viewed as a pair of two states pertaining to the constituent subsystems is called entangled. In the quantum world the whole is not the collection of the parts. With this discussion as background we can now give more formal account of the states of composite quantum systems.

2.2. **Tensor products.** A mathematical formalism behind quantum theory of composite systems is based on a tensor product (also known as a direct product). We use tensor products to construct Hilbert spaces associated with composed systems.



$\mathcal{H}_A \otimes \mathcal{H}_B$

Let states of system $\mathcal{A}$ be described by vectors in n dimensional Hilbert space $\mathcal{H}_A$ and states of system $\mathcal{B}$ by vectors in m dimensional Hilbert space $\mathcal{H}_B$. The combined system of $\mathcal{A}$ and $\mathcal{B}$ is then described by vectors in the nm dimensional tensor product space $\mathcal{H}_A \otimes \mathcal{H}_B$. Given bases $\{|a_1\rangle, \ldots, |a_n\rangle\}$ in $\mathcal{H}_A$ and $\{|b_1\rangle, \ldots, |b_m\rangle\}$ in $\mathcal{H}_B$ we form the tensor product basis consisting of the ordered pairs $|a_i\rangle \otimes |b_j\rangle$, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The tensor product space $\mathcal{H}_A \otimes \mathcal{H}_B$ consists of all linear combination of such tensor product basis vectors,

$$|\psi\rangle = \sum_{ij} c_{ij}\,|a_i\rangle \otimes |b_j\rangle\,.$$

The tensor product operation $\otimes$ is distributive:

$$\begin{aligned} |a\rangle \otimes (\beta_1\,|b_1\rangle + \beta_2\,|b_2\rangle) &= \beta_1\,|a\rangle \otimes |b_1\rangle + \beta_2\,|a\rangle \otimes |b_2\rangle\,, \\ (\alpha_1\,|a_1\rangle + \alpha_2\,|a_2\rangle) \otimes |b\rangle &= \alpha_1\,|a_1\rangle \otimes |b\rangle + \alpha_2\,|a_2\rangle \otimes |b\rangle\,. \end{aligned}$$

The tensor product of Hilbert spaces is a Hilbert space. The inner products on $\mathcal{H}_A$ and $\mathcal{H}_B$ give a natural inner product on $\mathcal{H}_A \otimes \mathcal{H}_B$. It is defined for any two product vectors

$$\left(\langle a'| \otimes \langle b'|\right)(|a\rangle \otimes |b\rangle) = \langle\, a' \,|\, a \,\rangle\, \langle\, b' \,|\, b \,\rangle \tag{4}$$

and extended by linearity to sums of tensor products of vectors and by associativity, $(\mathcal{H}_a \otimes \mathcal{H}_b) \otimes \mathcal{H}_c = \mathcal{H}_a \otimes (\mathcal{H}_b \otimes \mathcal{H}_c)$, to any number of subsystems. Some joint states of $\mathcal{A}$ and $\mathcal{B}$ can be expressed as a single tensor product, $|\psi\rangle = |a\rangle \otimes |b\rangle$ (often written as $|a\rangle\,|b\rangle$, or $|a, b\rangle$, or even $|ab\rangle$), meaning that subsystem $\mathcal{A}$ is in state $|a\rangle$ and subsystem $\mathcal{B}$ in state $|b\rangle$. If we expand $|a\rangle = \sum_i \alpha_i\,|a_i\rangle$ and $|b\rangle = \sum_i \beta_j\,|b_j\rangle$ then $|\psi\rangle = \sum_{ij} \alpha_i\beta_j\,|a_i\rangle \otimes |b_j\rangle$ and we see that for all such states the coefficients c_{ij} in Eq. (7) are of a rather special form $\alpha_i\beta_j$. We call such states separable (or just product states). All states which are not separable are called entangled.

The bra corresponding to the tensor product state $|a\rangle \otimes |b\rangle$ is written as $(|a\rangle \otimes |b\rangle)^\dagger = \langle a| \otimes \langle b|$ where the order of the factors on either side of $\otimes$ does not change when the dagger operation is applied.

If the bases $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$ are orthonormal then so is the tensor product basis $\{|a_i\rangle \otimes |b_j\rangle\}$.

We will also need the concept of the tensor product of two operators. If A is an operator on $\mathcal{H}_A$ and B an operator on $\mathcal{H}_B$ then the tensor product operator $A \otimes B$ is an operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ defined by its action on product vectors

$$(A \otimes B)\,(|a\rangle \otimes |b\rangle) = (A\,|a\rangle) \otimes (B\,|b\rangle).$$

Its action on all other vectors is determined by linearity,

$$A \otimes B \left(\sum_{ij} c_{ij} \left| a_i \right\rangle \otimes \left| b_j \right\rangle \right) = \sum_{ij} c_{ij} \, A \left| a_i \right\rangle \otimes B \left| b_j \right\rangle .$$

2.3. **Back to qubits.** Let us see how this formalism works for qubits. The n-fold tensor products of vectors from the standard basis $\{|0\rangle, |1\rangle\}$ represent binary strings of length n, e.g. for $n = 3$,

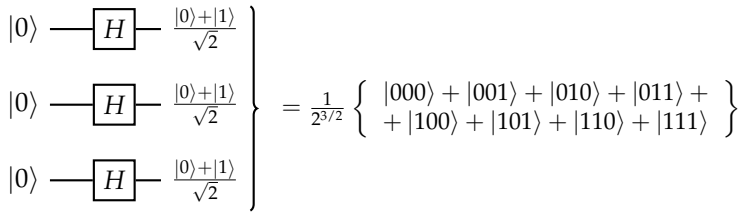
$$\begin{aligned} |0\rangle \otimes |1\rangle \otimes |1\rangle &\equiv |011\rangle, \\ |1\rangle \otimes |1\rangle \otimes |1\rangle &\equiv |111\rangle. \end{aligned}$$

We often drop the $\otimes$ symbol, especially when we deal with the standard tensor product basis. For example, a state of a quantum register composed of four qubits holding binary string 1001 may be is written as $|1\rangle \otimes |0\rangle \otimes |0\rangle \otimes |1\rangle$ or $|1\rangle\,|0\rangle\,|0\rangle\,|1\rangle$, or simply as $|1001\rangle$.

A classical register composed of three bits can store only one of these two binary strings at any time. A quantum register composed of three qubits can store both of them in a superposition. Indeed, if we start with the state $|011\rangle$ and apply the Hadamard gate to the first qubit ($H \otimes \mathbb{1} \otimes \mathbb{1}$) then, given that linear combinations distribute over the tensor products, we obtain

$$|011\rangle \longrightarrow \frac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle \right) \otimes |1\rangle \otimes |1\rangle \quad \equiv \quad \frac{1}{\sqrt{2}} \left(|011\rangle + |111\rangle \right). \tag{5}$$

In fact, we can prepare this register in a superposition of all eight binary strings it can hold. Let us apply the tensor product operation $H \otimes H \otimes H$, often written as $H^{\otimes 3}$, to the state $|0\rangle \otimes |0\rangle \otimes |0\rangle \equiv |000\rangle$,



The resulting state

$$\frac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle \right) \otimes \frac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle \right) \otimes \frac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle \right), \tag{6}$$

is a superposition of all binary strings of length 3,

$$\frac{1}{2^{3/2}} \sum_{x \in \{0,1\}^3} |x\rangle = \frac{1}{2^{3/2}} \left(|000\rangle + |001\rangle + |010\rangle + |011\rangle + |100\rangle + |101\rangle + |110\rangle + |111\rangle \right).$$

The tensor product operation $H^{\otimes n}$, meaning "apply the Hadamard gate to each of your n qubits" is known as the Hadamard Transform and, like the Hadamard gate in the quantum interference circuit, it opens and closes a multi-qubit interference. The Hadamard Transforms maps product states into product states.

2.4. **Separable or entangled.** Let us stress again that most vectors in $\mathcal{H}_a \otimes \mathcal{H}_b$ are entangled and cannot be written as product states $|a\rangle \otimes |b\rangle$ with $|a\rangle \in \mathcal{H}_a$ and $|b\rangle \in \mathcal{H}_b$. In order to see this let us write any joint state $|\psi\rangle$ of $\mathcal{A}$ and B in a product basis as

$$|\psi\rangle = \sum_{ij} c_{ij} \left| a_i \right\rangle \otimes \left| b_j \right\rangle = \sum_i |a_i\rangle \otimes \left(\sum_j c_{ij} \left| b_j \right\rangle \right) = \sum_i |a_i\rangle \otimes |\phi_i\rangle, \tag{7}$$

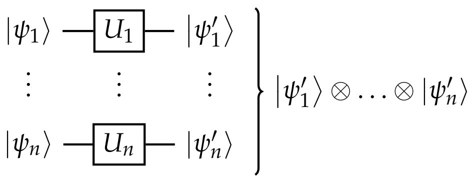
where the $|\phi_i\rangle = \sum_j c_{ij} \left| b_j \right\rangle$ are vectors in $\mathcal{H}_B$ that need not be normalised. For all product states these vectors have a special form. Indeed, if $|\psi\rangle = |a\rangle \otimes |b\rangle$ then, after expanding the first state in the $|a_i\rangle$ basis, we obtain

$$|\psi\rangle = \sum_i |a_i\rangle \otimes \left(\sum_i \alpha_i \left| b \right\rangle \right).$$

This expression has the same form as Eq. (7) with $|\phi_i\rangle = \alpha_i |b\rangle$, i.e. each of the $|\phi_i\rangle$ vectors in this expansion is a multiple of the same vector $|b\rangle$. Conversely, if $|\phi_i\rangle = \alpha_i |b\rangle$ for all i in Eq. (7) then $|\psi\rangle$ must be a product state. Thus if we want to identify which joint states are product states and which are not we simply write the joint state according to Eq. (7) and check if all the $|\phi_i\rangle$ vectors are multiples of a single vector. Needless to say, if we choose states $|\phi\rangle$ randomly it is very unlikely that this condition is satisfied and we almost certainly pick an entangled state. Quantum entanglement is one of the most fascinating aspects of quantum theory. We will now explore some of its implications.

Even though an entangled state cannot be written as a tensor product it can always be written as a linear combination of vectors from the tensor product basis. In fact, any state of n qubits $|\psi\rangle$ can be expressed in the standard *product* basis. In general, given n qubits we need $2(2^n - 1)$ real parameters to describe their state vector, but only $2n$ to describe separable states.

2.5. **Controlled-NOT.** How do entangled states arise in real physical situations? The short answer is that entanglement is the result of interactions. It is easy to see that tensor product operations, $U_1 \otimes \ldots \otimes U_n$, map product states into product states



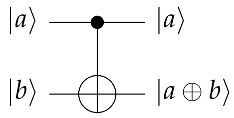
thus any collection of separable qubits remains separable. However, as soon as qubits start interacting with each other they become entangled and things get really interesting. We will describe interactions between qubits in terms of entangling gates. These are simple unitary operations that cannot be written as tensor products of unitary operations on individual qubits. The most popular two-qubit entangling gate is the controlled-NOT (C-NOT), also known as the controlled-X gate (here X refers to the Pauli $\sigma_x \equiv X$ operation that effects the bit-flip). The gate acts on two qubits — it flips the second (target) qubit if the first (control) qubit is $|1\rangle$ and does nothing if the control qubit is $|0\rangle$. It is represented by the unitary matrix written in the standard basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$,

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

C-NOT gate

$|a\rangle |b\rangle \mapsto |a\rangle |a \oplus b\rangle$

$$\text{C-NOT} = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right] \tag{8}$$



where $a, b = 0$ or 1 and $\oplus$ denotes XOR or addition modulo 2. We can write this operation as

$$|a\rangle |b\rangle \mapsto |a\rangle |a \oplus b\rangle .$$

The gate is described by the matrix that does not admit any tensor product decomposition, but it can be written as the sum of tensor products

$$|0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes X,$$

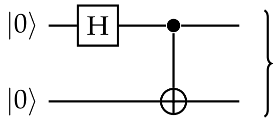
where X is the Pauli bit-flip operation. Let us now discuss all kind of interesting things that you can (and cannot) do with a controlled-NOT gate.

Make sure that you understand how the Dirac notation is used here. Think why the expression $|0\rangle\langle 0| \otimes A + |1\rangle\langle 1| \otimes B$ means "if the first qubit is in state $|0\rangle$ apply A to the second one but if the first qubit is in state $|1\rangle$ apply B to the second one". What happens if the first qubit in a superposition of $|0\rangle$ and $|1\rangle$?

2.6. **The Bell states and the Bell measurement.** Let us start with the generation of entanglement. Here is a simple circuit that demonstrates the entangling power of C-NOT

For whom the bell tolls

John Stewart Bell (1928–1990) was a Northern Irish physicist.



$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

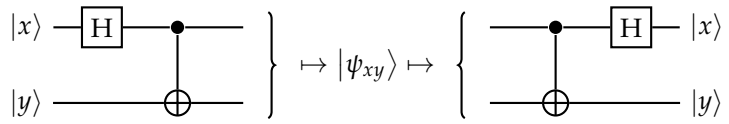
The separable input $|0\rangle\,|0\rangle$ evolves as

$$\begin{aligned} |0\rangle\,|0\rangle \xrightarrow{H} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\,|0\rangle \quad &= \quad \frac{1}{\sqrt{2}}\,|0\rangle\,|0\rangle + \frac{1}{\sqrt{2}}\,|1\rangle\,|0\rangle \\ &\xrightarrow{\text{C-NOT}} \quad \frac{1}{\sqrt{2}}\,|0\rangle\,|0\rangle + \frac{1}{\sqrt{2}}\,|1\rangle\,|1\rangle\,, \end{aligned} \tag{9}$$

resulting in the entangled output $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. In fact, this circuit implements the unitary operation which maps the standard computation basis into the four entangled states, known as the Bell states,

$$\begin{aligned} |00\rangle &\mapsto |\psi_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \\ |01\rangle &\mapsto |\psi_{01}\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\ |10\rangle &\mapsto |\psi_{10}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |11\rangle &\mapsto |\psi_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{aligned}$$

The Bell states form an orthonormal basis in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ of two qubits. We can perform measurements in the Bell basis. The easiest way to do it in practice is to "rotate" the Bell basis to the standard basis and then perform the measurement in the standard basis.

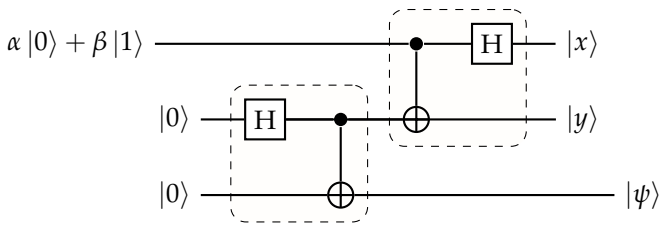


As we have just seen, the circuit on the left maps the standard basis $|x\rangle\,|y\rangle \equiv |xy\rangle$ into the four Bell states $|\psi_{xy}\rangle$. The circuit on the right, which is the reverse image of the circuit on the left, implements the inverse of this operation and maps the Bell states $|\psi_{xy}\rangle$ into the corresponding states from the standard basis $|xy\rangle$. This unitary mapping allows us to "implement" the projections on the Bell states by applying the circuit (on the right) followed by the regular qubit by qubit measurement in the standard basis.

For any state $|\psi\rangle$ of two qubits the amplitude $\langle\,\psi_{xy}\,|\,\psi\,\rangle$ can be written as $\langle xy|\,U^\dagger\,|\psi\rangle$, where $U^\dagger$ is the compensating unitary, such that $|\psi_{xy}\rangle = U\,|xy\rangle$.

2.7. **Quantum teleportation.** An unknown quantum state can be teleported from one location to another. Consider the following circuit

Divide et impera, that is, divide and conquer, a good approach to solving problems in mathematics (and in life). Start with smaller circuits, those surrounded by the dashed boxes.



The first input qubit (counting from the top) is in some arbitrary state. After the action of the circuit in the first dashed box the state of the three qubits reads (we have dropped the normalisation factors),

$$(\alpha\,|0\rangle + \beta\,|1\rangle)(|00\rangle + |11\rangle).$$

By regrouping the terms, but keeping the qubits in the same order, this state can be written as the sum

$$\begin{array}{lcl}(|00\rangle + |11\rangle) & \otimes & (\alpha\,|0\rangle + \beta\,|1\rangle) + \\ (|01\rangle + |10\rangle) & \otimes & (\alpha\,|1\rangle + \beta\,|0\rangle) + \\ (|00\rangle - |11\rangle) & \otimes & (\alpha\,|0\rangle - \beta\,|1\rangle) + \\ (|01\rangle - |10\rangle) & \otimes & (\alpha\,|1\rangle - \beta\,|0\rangle).\end{array} \tag{10}$$

The second dashed box circuit maps the four Bell states of qubits 1 and 2 to the corresponding states from the computational basis

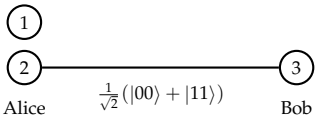
$$\begin{array}{lcl}|00\rangle & \otimes & (\alpha\,|0\rangle + \beta\,|1\rangle) + \\ |01\rangle & \otimes & (\alpha\,|1\rangle + \beta\,|0\rangle) + \\ |10\rangle & \otimes & (\alpha\,|0\rangle - \beta\,|1\rangle) + \\ |11\rangle & \otimes & (\alpha\,|1\rangle - \beta\,|0\rangle).\end{array} \tag{11}$$

Upon performing the standard measurement and learning the values of x and y we can choose one of the four transformations,

$$00 \to \mathbb{1}, \quad 01 \to X, \quad 10 \to Z, \quad 11 \to ZX, \tag{12}$$

(e.g. if $x = 0, y = 1$ we choose X) and apply it to the third qubit. This restores the original state of the first qubit. If you understand how this circuit works then you are ready for quantum teleportation. Here is a dramatic version.

2.8. **Saving Cambridge science.** Suppose three qubits, which look very similar, are initially in a possession of an absent-minded Oxford student Alice. The first qubit is in a precious quantum state and this state is needed urgently for an experiment in Cambridge. The other two qubits are entangled, in the $|\psi_{00}\rangle$ state. Alice's colleague, Bob, pops in to collect the qubit. Once he is gone Alice realises that by mistake she gave him not the first but the third qubit, the one which is entangled with the second qubit. The situation seems to be hopeless – Alice does not know the quantum state of the first qubit, Bob is now miles away and her communication with him is limited to few bits. However, Alice and Bob are both very clever and they both diligently attended the "Introduction to Quantum Information Science" course at Oxford. Can Alice rectify her mistake and save Cambridge science? Hmmm... pause for thought.... Sure she can. Alice can teleport the state of the first qubit. She performs the Bell measurement on the first two qubits, which gives her two binary digits, x and y. She then broadcasts x and y to Bob who chooses one of the four transformations, as in Eq. (12), and recovers the original state.



2.9. **Thou shalt not clone.** Let us now look at something that controlled-NOT seems to be doing but in fact it doesn't. It is easy to see that the c-NOT can copy the bit value of the first qubit,

$$|a\rangle|0\rangle \mapsto |a\rangle|a\rangle \qquad a = 0, 1. \tag{13}$$

One might suppose that this gate could also be used to copy superpositions such as $|\psi\rangle = \alpha_0\,|0\rangle + \alpha_1\,|1\rangle$, so that

$$|\psi\rangle|0\rangle \mapsto |\psi\rangle|\psi\rangle \tag{14}$$

for any $|\psi\rangle$. This is not so! The unitarity of the c-NOT requires that the gate turns superpositions in the control qubit into *entanglement* of the control and the target. If the control qubit is in a superposition state $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$, $(\alpha_0, \alpha_1 \neq 0)$, and the target in $|0\rangle$ then the c-NOT generates the entangled state

Quantum states cannot be cloned.

$$(\alpha|0\rangle + \alpha_1|1\rangle)\,|0\rangle \mapsto \alpha_0|00\rangle + \alpha_1|11\rangle. \tag{15}$$

In fact, it is impossible to clone an unknown quantum state. To see this assume that you can build a universal quantum cloner. Take any two normalised states $|\psi\rangle$ and

$|\phi\rangle$ which are non-identical ($|\langle\psi|\phi\rangle| \neq 1$) and non-orthogonal ($\langle\psi|\phi\rangle \neq 0$), and run your hypothetical cloning machine,

$$|\psi\rangle|0\rangle|W\rangle \quad \mapsto \quad |\psi\rangle|\psi\rangle|W'\rangle, \tag{16}$$

$$|\phi\rangle|0\rangle|W\rangle \quad \mapsto \quad |\phi\rangle|\phi\rangle|W''\rangle. \tag{17}$$

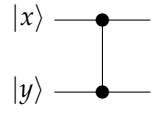
Here, the third system, initially in state $|W\rangle$, represents everything else (say, the internal state of the cloning machine). For this be a unitary transformation which preserves the inner product hence we must require

$$\langle\psi|\phi\rangle = \langle\psi|\phi\rangle^2\langle W'|W''\rangle, \tag{18}$$

which can only be satisfied when $|\langle\psi|\phi\rangle| = 0$ or 1, which contradicts our assumptions. Thus states of qubits, unlike states of classical bits, cannot be faithfully cloned. Note, that in quantum teleportation the original state must be destroyed for otherwise we would produce a clone of an unknown quantum state. The no-cloning property of quantum states leads to interesting applications, quantum cryptography being one such.

2.10. **Controlled-phase.** Needless to say, it is not all about the controlled-NOT gates. Another common two-qubit gate is the controlled phase shift gate cP_φ defined as

$$cP_\varphi = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\varphi} \end{array}\right] \qquad \left.\begin{array}{l} |x\rangle \\ |y\rangle \end{array}\right\} e^{ixy\varphi}\,|\,x\rangle\,|\,y\rangle\,. \tag{19}$$



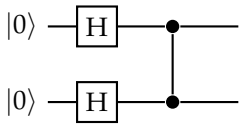
Again, the matrix is written in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ and the diagram on the right shows the structure of the gate. If we do not specify the phase we usually assume that $\varphi = \pi$, in which case we call this operation the controlled-Z gate,

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{array}\right]$$

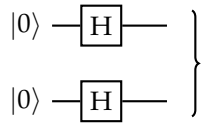
c-Z gate

$$|0\rangle\langle 0|\,0 \otimes \mathbb{1} + |1\rangle\langle 1|\,1 \otimes Z.$$

Here Z refers to the Pauli phase-flip, $\sigma_z \equiv Z$, operation. In order to see its entangling power consider the following circuit

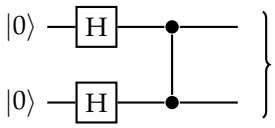


First, the two Hadamard gates prepare the equally weighted superposition of all states from the computational basis



$$\tfrac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$$

and then the controlled-Z operation flips the sign in front of $|11\rangle$,



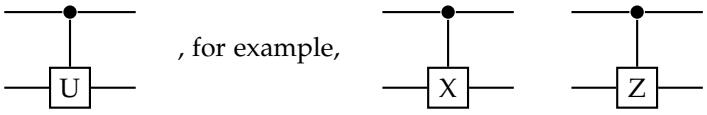
$$\tfrac{1}{2}(|00\rangle + |01\rangle + |10\rangle - |11\rangle)$$

which results in the entangled state (see exercises). As you can see, sometimes introducing a tiny relative phase shift can result in entangling two systems.

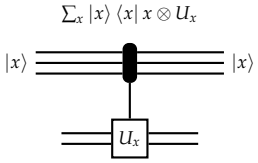
2.11. **Controlled-U.** More generally, these various 2-qubit controlled gates are all of the form controlled-U, for some single-qubit unitary transformation U,

$$|0\rangle\langle 0|\,0\otimes\mathbb{1}+|1\rangle\langle 1|\,1\otimes U.$$

It is graphically represented as



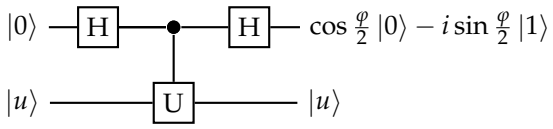
which is an alternative way of representing controlled-X (controlled-NOT) and controlled-Z gates respectively. We can go further and consider a more general unitary operation, namely, an x-controlled-U on two qubits,



$$\sum_x |x\rangle\langle x|\otimes U_x \equiv |0\rangle\langle 0|\otimes U_0+|1\rangle\langle 1|\otimes U_1,$$

where U_0 and U_1 are unitary transformation applied to the second qubit if the first one is in state $|0\rangle$ and $|1\rangle$ respectively. In general, an x-controlled-U is a unitary operation $\sum_x |x\rangle\langle x|\otimes U_x$ on two registers of size n and m; here $x\in\{0,1\}^n$ and U_x is the corresponding $2^m\times 2^m$ unitary matrix acting on the second register.

2.12. **Phase kickback.** Before we conclude this lecture, let me describe a simple "trick", an unusual way of introducing phase shifts, which will be essential for our analysis of quantum algorithms. Consider the following circuit



You recognise, I hope, the interference circuit at the top. Well, almost, instead of a phase gate I have inserted a controlled-U operation but, as you will see in a moment, it will mimic the phase gate. The second qubit is prepared in state $|u\rangle$ which is an eigenstate of U, that is, $U|u\rangle = e^{i\varphi}|u\rangle$. The circuit effects the following sequence of transformations (normalisation factors neglected)

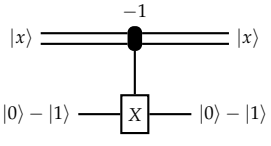
$$\begin{aligned}
|0\rangle|u\rangle &\overset{H}{\mapsto} (|0\rangle+|1\rangle)|u\rangle = |0\rangle|u\rangle+|1\rangle|u\rangle\\
&\overset{cU}{\mapsto} |0\rangle|u\rangle+|1\rangle U|u\rangle = |0\rangle|u\rangle+e^{i\varphi}|1\rangle|u\rangle = (|0\rangle+e^{i\varphi}|1\rangle)|u\rangle\\
&\overset{H}{\mapsto} \left(\cos\frac{\varphi}{2}|0\rangle - i\sin\frac{\varphi}{2}|1\rangle\right)|u\rangle
\end{aligned}$$

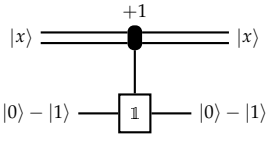
Note that the second qubit does not get entangled with the first one, it retains its original state $|u\rangle$. The interaction between the two qubits, induced by the controlled-U gate, introduces a phase shift on the first qubit. This may look like an unnecessarily complicated way of introducing phase shifts, but, as we shall see soon, this is how quantum computers do it. Let me give you a preview of things to come. Consider the following x-controlled-U operation,

$$|00\rangle\langle 00|\,00\otimes\mathbb{1}+|01\rangle\langle 01|\,01\otimes\mathbb{1}+|10\rangle\langle 10|\,10\otimes\mathbb{1}+|11\rangle\langle 11|\,11\otimes X. \qquad (20)$$

$$\left[\begin{array}{c|c|c|c} \mathbb{1} & 0 & 0 & 0\\ \hline 0 & \mathbb{1} & 0 & 0\\ \hline 0 & 0 & \mathbb{1} & 0\\ \hline 0 & 0 & 0 & X \end{array}\right]$$

The corresponding matrix is shown in the margin. The first register is of size 2 and the second register is of size 1 (just a single qubit). The expression above tells you that if the first register is prepared in state $|11\rangle$ then the qubit in the second register is flipped (the Pauli bit-flip X operation is applied to the second register) and nothing happens otherwise (the identity $\mathbb{1}$ is applied to the second register whenever the first register is in state $|00\rangle, |01\rangle$ or $|10\rangle$). This unitary operation is a quantum version of the Boolean function evaluation, it corresponds to the Boolean function $f:\{0,1\}^2\mapsto\{0,1\}$ such that $f(11)=1$ and $f(00)=f(01)=f(10)=0$.





Whenever $f(x) = 1$ we flip the bit value in the second register (with operation X) and whenever $f(x) = 0$ we do nothing. Now, prepare the qubit in the second register in state $|0\rangle - |1\rangle$. This is the eigenstate of X with eigenvalue -1. Thus, whenever X is applied to the second register the phase factor -1 appears in front of the corresponding term in the first register. If we prepare the first register in the superposition $|00\rangle + |01\rangle + |10\rangle + |11\rangle$ then the result of applying the x-controlled-U, given by (20), is the entangled state $|00\rangle + |01\rangle + |10\rangle - |11\rangle$. The phase kickback mechanism introduced a relative phase in the equally weighted superposition of all binary strings of size two. This is how we control quantum interference in quantum computation. We will return to this topic in our next lecture, when we discuss quantum evaluation of Boolean functions and quantum algorithms.

2.13. **Universality revisited.** We will come across few more gates in this course but at this stage you already know all the elementary unitary operations that are needed to construct any unitary operation on any number of qubits. The Hadamard gate, all phase gates, and the C-NOT, form a *universal set of gates* i.e. if the C-NOT gate as well as the Hadamard and all phase gates are available then any n-qubit unitary operation can be constructed exactly with $O(4^n n)$ such gates. We should mention that there are many universal sets of gates. In fact, almost any gate which can entangle two qubits can be used as a universal gate. We will be in particular interested in a *finite* universal set of gates, such as the one containing the Hadamard, $P_{\frac{\pi}{4}}$ (the T gate) and the C-NOT, can approximate any unitary operation on n qubits with arbitrary precision. The price to pay is the number of gates — better precision requires more gates. We shall elaborate on it later on.

Here and in the following we use asymptotic notation: given a positive function $f(n)$, the symbol $O(f(n))$ means bounded from above by $c\,f(n)$ for some constant $c > 0$ (for sufficiently large n). For example, $15n^2 + 4n + 7$ is $O(n^2)$.

2.14. **Spoiler: density operators and the like.** The existence of entangled states begs an obvious question: if we cannot attribute a state vectors to an individual qubit then how shall we describe its quantum state? In the next few lectures we will see that when we limit our attention to a part of a larger system, then states are not represented by vectors, measurements are not described by orthogonal projections and evolution is not unitary. As a spoiler, here is a list of few new concepts that will be introduced soon.

$$\begin{aligned} \text{state vectors} &\mapsto \text{density operators} \\ \text{unitary evolution} &\mapsto \text{completely positive trace preserving map} \\ \text{orthogonal projectors} &\mapsto \text{positive operator-valued measure} \end{aligned}$$

Self guided tour: tensor products in components

In our discussion of tensor products we have taken a rather abstract approach. There are, however, situations in which we have to put numbers in and write tensor products of vectors and matrices explicitly. For example, here is the standard basis of two qubits written explicitly as column vectors,

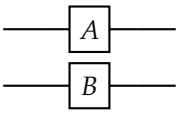
$$|00\rangle \equiv |0\rangle \otimes |0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad , \qquad |01\rangle \equiv |0\rangle \otimes |1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$|10\rangle \equiv |1\rangle \otimes |0\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad , \qquad |11\rangle \equiv |1\rangle \otimes |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Given $|a\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle$ and $|b\rangle = \beta_0 |0\rangle + \beta_1 |1\rangle$ we write $|a\rangle \otimes |b\rangle$ as

$$|a\rangle \otimes |b\rangle = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \otimes \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \alpha_0 \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \\ \alpha_1 \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \alpha_0\beta_0 \\ \alpha_0\beta_1 \\ \alpha_1\beta_0 \\ \alpha_1\beta_1 \end{bmatrix}.$$

Note that each element of the first vector multiplies the entire second vector. This is often the easiest way to get the tensor products in practice. The matrix elements of the tensor product operation $A \otimes B$



are given by

$$(A \otimes B)_{ik,jl} = A_{ij} B_{kl}. \tag{21}$$

In practice we just form block diagonal matrices, e.g.,

$$A \otimes B = \begin{bmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{bmatrix} \otimes \begin{bmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{bmatrix} = \left[\begin{array}{c|c} A_{00}B & A_{01}B \\ \hline A_{10}B & A_{11}B \end{array}\right],$$

where each element of the first matrix multiplies the entire second matrix,

$$\left[\begin{array}{c|c} A_{00}B & A_{01}B \\ \hline A_{10}B & A_{11}B \end{array}\right] = \left[\begin{array}{cc|cc} A_{00}B_{00} & A_{00}B_{01} & A_{01}B_{00} & A_{01}B_{01} \\ A_{00}B_{10} & A_{00}B_{11} & A_{01}B_{10} & A_{01}B_{11} \\ \hline A_{10}B_{00} & A_{10}B_{01} & A_{11}B_{00} & A_{11}B_{01} \\ A_{10}B_{10} & A_{10}B_{11} & A_{11}B_{10} & A_{11}B_{11} \end{array}\right]. \tag{22}$$

The tensor product matrix has composite indices, $(A \otimes B)_{ik,jl}$, here $ik = 00, 01, 10, 11$ labels rows and $jl = 00, 01, 10, 11$ labels columns and we always use the lexicographical order, $00, 01, 10, 11$. For example, as you can see above, $(A \otimes B)_{01,11}$ is the entry in the second row and the fourth column and reads $A_{01}B_{11}$.

Tensor product induces natural partition of matrices into blocks. Multiplication of block matrices works pretty much the same as regular matrix multiplication (assuming the dimensions of the sub matrices are appropriate), except that the entries are now matrices rather than numbers and they may not commute.

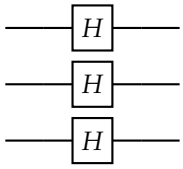
(1) Evaluate the matrix product of the following 4×4 block matrices (X, Y, Z are the Pauli matrices)

$$\left[\begin{array}{c|c} \mathbb{1} & X \\ \hline Y & Z \end{array}\right] \left[\begin{array}{c|c} \mathbb{1} & Y \\ \hline X & Z \end{array}\right] \tag{23}$$

Use the matrix form of $A \otimes B$ in Eq. (22) and explain how the following operations are performed on the block matrix

(2) Transposition $(A \otimes B)^T$ and partial transpositions $A^T \otimes B$, $A \otimes B^T$.
(3) Trace $\mathrm{Tr}\,(A \otimes B)$ and partial traces $(\mathrm{Tr}\,A) \otimes B$, $A \otimes (\mathrm{Tr}\,B)$.

Let us consider the tensor product of the Hadamard gates, the Hadamard Transform on three qubits,



This tensor product operation $H \otimes H \otimes H$, also written as $H^{\otimes 3}$, is described by a $2^3 \times 2^3$ matrix. We start with H and evaluate $H \otimes H$,

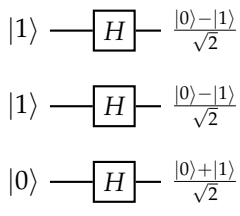
$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad H \otimes H = \frac{1}{2} \left[\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \hline 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{array}\right],$$

and once we have $H \otimes H$ we can tensor it again, $(H \otimes H) \otimes H = H \otimes H \otimes H$,

$$H \otimes H \otimes H = \left(\frac{1}{2}\right)^{\frac{3}{2}} \left[\begin{array}{cc|cc|cc|cc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ \hline 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ \hline 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ \hline 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{array}\right].$$

Can you find the entry $(H \otimes H \otimes H)_{010,101}$?

The rows and columns of $H \otimes H \otimes H$ are labelled by the triplets $000, 001, \ldots, 111$. Now, suppose that we act with $H^{\otimes 3}$ on state $|110\rangle$,



$$= \frac{1}{2^{3/2}} \left\{ \begin{array}{l} |000\rangle + |001\rangle - |010\rangle - |011\rangle + \\ -|100\rangle - |101\rangle + |110\rangle + |111\rangle \end{array} \right\}$$

(4) The output state is a superposition of all binary strings, $\sum_x c_x |x\rangle$, with $x \in \{0,1\}^3$. Where in the $H^{\otimes 3}$ matrix will you find the coefficients c_x?

Want to write down $H \otimes H \otimes H \otimes H$? I don't think so. This is an exponentially growing monster and you may soon run out of space if you really want to write it down. Instead spot the pattern of entries ± 1 in these matrices.

(5) Consider the Hadamard gate matrix H_{ab}, where $a, b = 0, 1$ are the labels for the rows and the columns. Observe that $H_{ab} = (-1)^{ab}/\sqrt{2}$. This may look like a fancy way of writing the entries of the Hadamard matrix but it will pay off in a moment. Use Eq. (21), or any other method, and analyse the pattern of ± 1s in the tensor product of Hadamard matrices. What is the entry $H^{\otimes 4}_{0101,1110}$?
(6) Show that up to the constant $(1/\sqrt{2})^n$ the entry $H^{\otimes n}_{a,b}$, for any n and for any binary strings a and b of length n, is $(-1)^{a \cdot b}$.

For any two binary strings of the same length $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ we can define their "scalar" the product as $a \cdot b = (a_1 b_1 \oplus \ldots \oplus a_n b_n)$.

(7) Show that $H^{\otimes n}$ maps

$$|a\rangle \mapsto \left(\frac{1}{\sqrt{2}}\right)^n \sum_{b\in\{0,1\}^n} (-1)^{a\cdot b} |b\rangle$$

(8) A quantum register of 10 qubits holds the binary string 0110101001. The Hadamard Transform is then applied to this register yielding a superposition of all binary strings of length 10. What is the sign in front of the $|0101010101\rangle$ term?

Self guided tour: the Schmidt decomposition

An arbitrary vector in the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ can be expanded in a product basis as

$$|\psi\rangle = \sum_{ij} c_{ij} |a_i\rangle |b_j\rangle ,$$

Moreover, for each particular joint state $|\psi\rangle$ we can find orthonormal bases, $\{|\tilde{a}_i\rangle\}$ in $\mathcal{H}_A$ and $\{|\tilde{b}_j\rangle\}$ in $\mathcal{H}_B$ such that $|\psi\rangle$ can be expressed as

$$|\psi\rangle = \sum_{i} d_i |\tilde{a}_i\rangle |\tilde{b}_i\rangle ,$$

where the coefficients d_i are nonnegative numbers. This is known as the Schmidt decomposition of $|\psi\rangle$. Any bipartite state can be expressed in this form but remember that the bases used depend on the state being expanded. Given two bipartite states $|\psi\rangle$ and $|\phi\rangle$ we usually cannot perform the Schmidt decomposition using the same orthonormal bases in $\mathcal{H}_A$ and $\mathcal{H}_B$. The number of terms in the Schmidt decomposition is, at most, the minimum of $\dim \mathcal{H}_A$ and $\dim \mathcal{H}_B$.

The Schmidt decomposition follows from the singular value decomposition (SVD); *any* $n \times m$ matrix C can be written as

$$C = UDV \tag{24}$$

where U and V are respectively $n \times n$ and $m \times m$ unitary matrices and D is an $n \times m$ diagonal matrix with real, non-negative elements in descending order $d_1 \geq d_2 \geq \ldots \geq d_{\min(n,m)}$ (the rest of the matrix is patched with zeros). The elements d_k are called singular values of C. You can visualize SVD if you view C as representing a linear transformation from m to n dimensional Euclidean space. It maps the unit ball in the m dimensional space to an ellipsoid in the n dimensional space. The singular values are the lengths of the semi-axes of that ellipsoid. The matrices U and V carry information about the locations of those axes and the vectors in the first space which map into them. Thus SVD tells us that the transformation C is composed of rotating the unit ball (transformation V), stretching the axes by factors d_k, and then rotating the resulting ellipsoid (transformation U).

Using the index notation $C_{ij} = \sum_k U_{ik} d_k V_{kj}$, thus we can apply SVD to c_{ij},

$$|\psi\rangle = \sum_{i,j} c_{ij} \ |a_i, b_j\rangle = \sum_{i,j} \sum_{k} U_{ik} d_k V_{kj} \ |a_i, b_j\rangle = \sum_{k} d_k \left(\sum_i U_{ik} |a_i\rangle \right) \otimes \left(\sum_j V_{kj} |b_j\rangle \right) .$$

The Schmidt decomposition of a separable state of the form $|a\rangle \otimes |b\rangle$ is trivially just this state. The Bell states Ψ^+ and Φ^+ are already written in their Schmidt form, whereas Ψ^- and Φ^- can be easily expressed in the Schmidt form, for example $|\Psi^-\rangle$ we have $d_1 = d_2 = \frac{1}{\sqrt{2}}$ and the Schmidt basis $|\bar{a}_1\rangle = |0\rangle , |\bar{a}_2\rangle = |1\rangle$, $|\bar{b}_1\rangle = |1\rangle , |\bar{b}_2\rangle = -|0\rangle$. The number of non-zero singular values of c_{ij} is called the rank of c_{ij}, or the rank of the corresponding quantum state, or sometimes, the Schmidt number. Clearly all bipartite states of rank one are separable. The Schmidt decomposition is almost unique. The ambiguity arises when we have two or more identical singular values, as, for example, in the case of the Bell states. Then any unitary transformation of the basis vectors corresponding to a degenerate singular value, both in $\mathcal{H}_a$ and in $\mathcal{H}_b$, generates another basis vectors.

Notes & Exercises

(1) **Entangled or not?** Let a joint state of $\mathcal{A}$ and $\mathcal{B}$ be written in a product basis as

$$|\psi\rangle = \sum_{ij} c_{ij} |a_i\rangle \otimes |b_j\rangle .$$

Assume that $\mathcal{H}_a$ and in $\mathcal{H}_b$ are of the same dimension. Show that if $|\psi\rangle$ is a product state then $\det c_{ij} = 0$. Show that the converse,

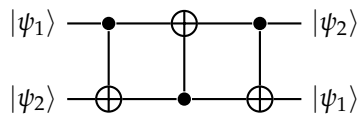
$$\det c_{ij} = 0 \Rightarrow |\psi\rangle \text{ is a product state}$$

holds only for qubits. Explain why. Deduce that the state $\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + (-1)^k |11\rangle)$ is entangled for $k = 1$ and unentangled for $k = 0$. Express the latter case explicitly as a product state.

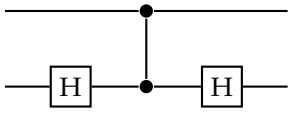
(2) There is lots of interesting physics behind this innocuous mathematical statement. For example, think again about the state $(|00\rangle + |11\rangle)/\sqrt{2}$. What happens if you measure just the first qubit? It is equally likely that you get $|0\rangle$ or $|1\rangle$, right? But after your measurement the two qubits are either in state $|00\rangle$ or in $|11\rangle$, i.e. they show the same bit value. Now, why might that be disturbing? Imagine the second qubit to be light-years away from the first one. It seems that that the measurement of the first qubit affects the second qubit right away, and that implies faster-than-light communication! This is what Einstein called "spooky action at a distance". But can you actually use this effect to send a message faster than light? What would happen if you tried? I hope you can see that it would not work, for the result of the measurement is random — you cannot choose the bit value you want to send. We shall return to this and related phenomena later on.

Spooky action at a distance is a loose translation of the German "spukhafte Fernwirkung", the term Albert Einstein used in his 1947 letter to Max Born.

(3) **Swap** Show that for any states $|\psi_1\rangle$ and $|\psi_2\rangle$ the circuit below effects the swap operation: $|\psi_1\rangle |\psi_2\rangle \mapsto |\psi_2\rangle |\psi_1\rangle$.



(4) Show that the circuit

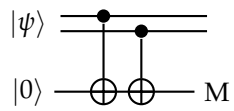


effects the controlled-NOT gate.

(5) The controlled-NOT gate can act as the measurement gate. If you prepare the target in state $|0\rangle$ the gate maps $|x\rangle |0\rangle \mapsto |x\rangle |x\rangle$, thus the target learns the bit value of the control qubit; it acts as a measuring device. If you wish, you can think about a subsequent measurement of the target qubit in the computational basis and and an observer learning about the bit value of the control qubit. Take a look at the circuit below. Here M stands for the measurement in the standard basis. Assume that the two top qubits are in the state

$$\frac{1}{\sqrt{3}} (|01\rangle - |10\rangle + i |11\rangle)$$

The measurement gives two outcomes, 0 and 1. What are the probabilities of the two outcomes and what is the post-measurement state in each case?



What is actually measured here?

(6) **Arbitrary controlled-U on two qubits** Any unitary operation U on a single qubit can be expressed as

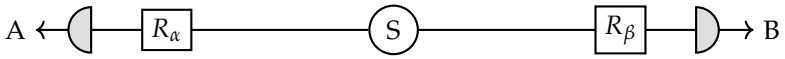
$$U = B^\dagger X B A^\dagger X A,$$

where X is the Pauli σ_x bit-flip operator and A and B are some unitaries. Suppose you can implement any single qubits gate and you have a couple of controlled-NOT gates. How would you implement any controlled-U operation on two qubits?

(7) **Entangled qubits.** Two entangled qubits in state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ are generated by source S; one qubit is sent to Alice and one to Bob, who perform measurements in the computational basis.

(a) What is the probability that Alice and Bob will register identical results? Can any correlations they observe be used for instantaneous communication?

(b) Prior to the measurements in the computational basis Alice and Bob apply unitary operations R_α and R_β to their respective qubits



The gate R_θ is defined by its action on the basis states

$$\begin{aligned} |0\rangle &\rightarrow \cos\theta\,|0\rangle + \sin\theta\,|1\rangle\,, \\ |1\rangle &\rightarrow -\sin\theta\,|0\rangle + \cos\theta\,|1\rangle\,. \end{aligned}$$

Show that the state of the two qubits prior to the measurements is

$$\cos(\alpha-\beta)\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) - \sin(\alpha-\beta)\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle).$$

What is the probability that Alice and Bob's outcomes are identical?

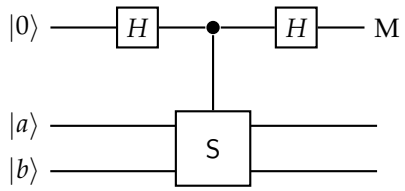
(8) **Quantum dense coding** (to be added)

(9) **Playing with conditional unitaries** The swap gate S on two qubits is defined first on product vectors, $S : |a\rangle\,|b\rangle \mapsto |b\rangle\,|a\rangle$ and then extended to sums of products vectors by linearity.

(a) Show that the four Bell states $\frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle)$, $\frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle)$ are eigenvectors of S which form the orthonormal basis in the Hilbert space associated with two qubits. Which Bell states span the symmetric subspace (all eigenvectors of S with eigenvalue 1) and which the antisymmetric one (all eigenvectors of S with eigenvalue -1)? Can S have any other eigenvalues except ± 1?

(b) Show that $P_\pm = \frac{1}{2}(\mathbb{1} \pm S)$ are two orthogonal projectors which form the decomposition of the identity and project on the symmetric and the antisymmetric subspaces. Decompose the state vector $|a\rangle\,|b\rangle$ of two qubits into symmetric and antisymmetric components.

(c) Consider the following quantum network composed of the two Hadamard gates, one controlled-S operation (also known as the controlled-swap or the Fredkin gate) and the measurement M in the computational basis,



The state vectors $|a\rangle$ and $|b\rangle$ are normalised but not orthogonal to each other. Step through the execution of this network, writing down quantum states of the three qubits after each computational step. What are the probabilities of observing 0 or 1 when the measurement M is performed?

(d) Explain why this quantum network implements projections on the symmetric and the antisymmetric subspaces of the two qubits.

(e) Two qubits are transmitted through a quantum channel which applies the same, randomly chosen, unitary operation U to each of them. Show that $U \otimes U$ leaves the symmetric and antisymmetric subspaces invariant.

(f) Polarised photons are transmitted through an optical fibre. Due to the variation of the refractive index along the fibre the polarisation of each photon is rotated by the same unknown angle. This makes communication based on polarisation encoding unreliable. However, if you can prepare any polarisation state of two photons you can still use the channel and communicate without any errors. How can this be achieved?

Complement 1

Why qubits? Why subsystems? Why entanglement?

One question that I hear over and over again is this: if entanglement is so fragile and so difficult to control then why bother, why not perform the whole computation in one physical system with sufficiently many quantum states which we can label in the same way we label states of qubits and give them the same computational meaning? It will not work for this is a very inefficient way of representing data (unary encoding). For serious computations we do need subsystems. Here is why.

Suppose you have n physical objects and each object has k distinguishable states. If you can access each object *separately* and put it into any of the k states then with only n operations you can prepare any of the k^n different configurations of the combined systems. Without any loss of generality let us put $k = 2$ and refer to each object of this type as a physical bit. We label the two states of the physical bit as 0 and 1. Any collection of n physical bits can be prepared in 2^n different configurations which can be used to store up to 2^n messages, or binary strings or 2^n different numbers. In order to represent numbers from 0 to $N-1$ we just have to choose n such that $N \leq 2^n$. Suppose the two states in the physical bit are separated by the energy difference ΔE then a preparation of any particular configuration will cost not more than $E = n\Delta E$ or $\log N \ \Delta E$ units of energy (the log is taken to the base 2).

In contrast if we choose to encode N configurations into one chunk of matter, say into the first N energy states of a single harmonic oscillator with the energy separation ΔE then, in the worst case, one has to use $E = N \ \Delta E$ units of energy (e.g. to go from the ground state labelled as 0 to the most excited state labelled as N). For large N this gives an exponieential gap in the energy expenditure between the binary encoding using physical bits and the so-called unary encoding using energy levels of harmonic oscillators.

One can, of course, try to switch from harmonic oscillators to quantum systems which have a finite spread in the energy spectrum. For example, by operating on the energy states of the hydrogen atom one can encode any number from 0 to $N-1$ and one is guaranteed not to spend more than $E_{max} = 13.6$ eV (otherwise the hydrogen atom is ionised). The snag is that in this case some of the electronic states will be separated by the energy difference of the order of E_{max}/N and to drive the system selectively from one state to another one has to tune into the frequency $E_{max}/\hbar N$ which requires a sufficiently long wavepacket (so that the frequency is well defined) and consequently the interaction time of the order $N(\hbar/E_{max})$. Thus we have to trade energy for time. It turns out that whichever way we try to represent the number N using the unary encoding, i.e. using N different states of a single chunk of matter, we end up depleting our physical resources, such as energy, time, space, at a much greater rate than in the case when we use subsystems. This plausibility argument indicates that for efficient processing of information the system must be divided into subsystems, for example, into physical bits.

MEASUREMENTS

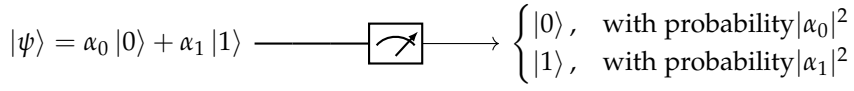
ARTUR EKERT

About quantum measurements, for, at some point, someone has to look at a measuring device and register the outcome. It turns out that this is a bit more tricky than one might think. Quantum measurement is not a passive acquisition of information; you measure you disturb. Even though it is a physical process, like any other quantum evolution, it is traditionally described by a different set of mathematical tools.

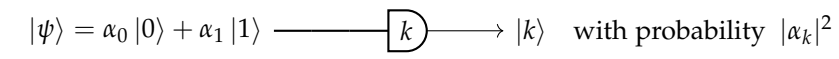
A mathematical setting for the quantum formalism is a Hilbert space $\mathcal{H}$, that is, a vector space with an inner product. The result of any preparation is represented by some unit vector $|\psi\rangle \in \mathcal{H}$ and any test is represented by some other unit vector $|e\rangle \in \mathcal{H}$. The inner product of these two vectors, $\langle e | \psi \rangle$, gives the probability amplitude that an object prepared in state $|\psi\rangle$ will pass a test for being in state $|e\rangle$. Probabilities are obtained by squaring absolute values of probability amplitudes, $|\langle e | \psi \rangle|^2 = \langle \psi | e \rangle \langle e | \psi \rangle$. After the test, in which the object was found to be in state $|e\rangle$, the object forgets about its previous state $|\psi\rangle$ and is indeed in state $|e\rangle$. This is the mysterious "quantum collapse", status of which we will discuss briefly later on. A more complete test involves states which form an orthonormal basis $\{|e_1\rangle, \ldots, |e_n\rangle\}$. These states are perfectly distinguishable from each other; the condition $\langle e_k | e_l \rangle = \delta_{kl}$ implies that a quantum system prepared in state $|e_l\rangle$ will never be found in state $|e_k\rangle$, if $k \neq l$. The probability amplitude that the system in state $|\psi\rangle$ will be found in state $|e_k\rangle$ is $\langle e_k | \psi \rangle$ and, given that vectors $|e_k\rangle$ span the whole vectors space, the system will be always found in one of the basis states, hence $\sum_k |\langle e_k | \psi \rangle|^2 = 1$. As a result, a complete measurement in quantum theory is determined by the choice of an orthonormal basis $\{|e_i\rangle\}$ in $\mathcal{H}$, and every such basis in principle represents a possible measurement.

The term "Hilbert space" used to be reserved for an infinite-dimensional inner product space that is complete i.e. every Cauchy sequence in the space converges to an element in the space. Nowadays, as in these notes, the term includes finite-dimensional spaces, which automatically satisfy the condition of completeness.

2.1. **Back to qubits.** The most common measurement in quantum information science is the standard measurement on a qubit, also referred to as the measurement in the standard or computational basis $\{|0\rangle, |1\rangle\}$. When we draw circuit diagrams it is tacitly assumed that such a measurement is performed on each qubit at the end of quantum evolution. However, if we want to emphasise the role of the measurement we can include it explicitly in the diagram as a special quantum gate, for example,



or, in an alternative notation, as

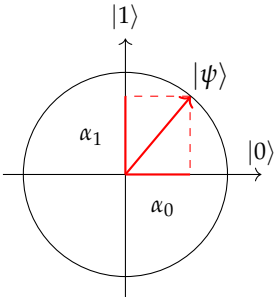


where $k = 0, 1$. As we can see from the diagrams, if the qubit is prepared in state $|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle$ and subsequently measured in the standard basis state the outcomes is $|0\rangle$ with probability

$$|\alpha_0|^2 = |\langle 0 | \psi \rangle|^2 = \underbrace{\langle \psi | 0 \rangle}_{\alpha_0^\star} \underbrace{\langle 0 | \psi \rangle}_{\alpha_0} = \langle \psi | \underbrace{|0\rangle \langle 0|}_{\text{projector}} | \psi \rangle = \langle \psi | P_0 | \psi \rangle ,$$

and $|1\rangle$ with probability

$$|\alpha_1|^2 = |\langle 1 | \psi \rangle|^2 = \underbrace{\langle \psi | 1 \rangle}_{\alpha_1^\star} \underbrace{\langle 1 | \psi \rangle}_{\alpha_1} = \langle \psi | \underbrace{|1\rangle \langle 1|}_{\text{projector}} | \psi \rangle = \langle \psi | P_1 | \psi \rangle ,$$



$|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle$

The standard, computational, basis defines the standard measurement.

Projectors

A projector, P, is any Hermitian operator, $P = P^\dagger$, which is idempotent, $P^2 = P$. The rank of P is evaluated using $\mathrm{Tr}(P)$. In the Dirac notation $|e\rangle \langle e|$ is a rank one projector on the subspace spanned by a unit vector $|e\rangle$. It acts on any vector $|v\rangle$ as $(|e\rangle \langle e|) |v\rangle = |e\rangle \langle e | v \rangle$.

where $P_0 = |0\rangle\langle 0|$ and $P_1 = |1\rangle\langle 1|$ are projectors on vectors $|0\rangle$ and $|1\rangle$ respectively. If the outcome of the measurement is $k = 0, 1$, the output state of the measurement gate is $|k\rangle$. The original state $|\psi\rangle$ is irretrievably lost. This sudden change of the state, from the pre-measurement state $|\psi\rangle$ to the post-measurement state, either $|0\rangle$ or $|1\rangle$ is often called a "collapse" or a "reduction" of the state.

It looks like there are two distinct ways for a quantum state to change. On the one hand we have unitary evolutions and on the other hand we have an abrupt change during the measurement process. Surely, the measurement process is not governed by any different laws of physics? It is not! A measurement is a physical process and can be explained without any "collapse" , but, it is usually a complicated process in which one complex system (a measuring apparatus or an observer) interacts and gets correlated with a physical system being measured. We will discuss it later on, but for now let us accept a "collapse" as a convenient mathematical shortcut and describe it in terms of projectors rather than unitary operators.

2.2. **The projection rule.** So far we have identified measurements with orthonormal bases, or, if you wish, with a set of orthonormal projectors on the basis vectors.

$$\langle e_k | e_l \rangle = \delta_{kl}.$$

The orthonormality condition. The basis consists of unit vectors which are pairwise orthogonal.

$$\sum_k |e_k\rangle\langle e_k| = \mathbb{1}.$$

The completeness condition means that *any* vector in $\mathcal{H}$ can be expressed as the sum of orthogonal projections on $|e_k\rangle$.

Given a quantum system in state $|\psi\rangle$, such that $|\psi\rangle = \sum_k \alpha_k |e_k\rangle$,

$$|\psi\rangle = \mathbb{1}|\psi\rangle = \sum_k |e_k\rangle\langle e_k|\psi\rangle = \sum_k |e_k\rangle\langle e_k|\psi\rangle = \sum_k |e_k\rangle\alpha_k = \sum_k \alpha_k |e_k\rangle,$$

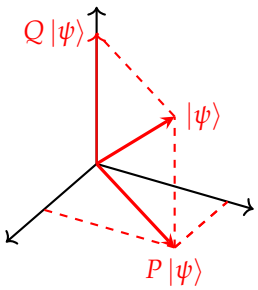
the measurement in the basis $\{|e_i\rangle\}$ gives outcome labelled by e_k with probability $|\langle e_k|\psi\rangle|^2 = \langle\psi|e_k\rangle\langle e_k|\psi\rangle$ and leaves the system in state $|e_k\rangle$. This is a complete measurement, which represents the best we can do in terms of resolving state vectors in the basis states. But sometimes we do not want our measurement to distinguish all the elements of an orthonormal basis. For example, a complete measurement in a four dimensional Hilbert space will have four distinct outcomes, $|e_1\rangle, |e_2\rangle, |e_3\rangle$, and $|e_4\rangle$, but we may want to lump together some of the outcomes and distinguish, say, only between $\{|e_1\rangle, |e_2\rangle\}$ and $\{|e_3\rangle, |e_4\rangle\}$. In other words, we might be trying to distinguish one subspace from another, without separating vectors that lie in the same subspace. Such measurements are possible and they can be less disruptive than the complete measurements. Intuitively, an incomplete measurement has fewer outcomes and hence is less informative but the state after such a measurement is usually less disturbed. In general, instead of projecting on one dimensional subspaces spanned by vectors from an orthonormal basis we can decompose our Hilbert space into mutually orthogonal subspaces of various dimensions and project on them.

$$P_k P_l = P_k \delta_{kl}$$

Orthogonality condition for projectors.

$$\sum_k P_k = \mathbb{1}$$

Decomposition of the identity



For any decomposition of the identity into orthogonal projectors P_k there exists a measurement that takes a quantum system in state $|\psi\rangle$, outputs label k, with probability $\langle\psi|P_k|\psi\rangle$ and leaves the system in the state $P_k|\psi\rangle$ (multiplied by the normalisation factor i.e. divided by the length of $P_k|\psi\rangle$),

$$|\psi\rangle \longrightarrow \frac{P_k|\psi\rangle}{\sqrt{\langle\psi|P_k|\psi\rangle}}.$$

2.3. **Example of an incomplete measurement.** Consider a three dimensional Hilbert space and the following two orthogonal projectors $P = |e_1\rangle\langle e_1| + |e_2\rangle\langle e_2|$ and $Q = |e_3\rangle\langle e_3|$ that form the decomposition of the identity, $P + Q = \mathbb{1}$. Suppose that a physical system is prepared in state $|\psi\rangle = \alpha_1|e_1\rangle + \alpha_2|e_2\rangle + \alpha_3|e_3\rangle$. Ideally we would like to perform a complete measurement that would resolve the state $|v\rangle$ into the three basis states but suppose our experimental apparatus is not good enough and lumps together $|e_1\rangle$ and $|e_2\rangle$. It can only differentiate between the two subspaces associated with projectors P and Q. The apparatus, in this incomplete measurement, may find the system in the subspace associated with P. This happens with the probability $\langle\psi|P|\psi\rangle = \langle\psi|e_1\rangle\langle e_1|\psi\rangle + \langle\psi|e_2\rangle\langle e_2|\psi\rangle = |\alpha_1|^2 + |\alpha_2|^2$, and the state right after the measurement is the normalised $P|\psi\rangle$, that is,

$$\frac{\alpha_1|1\rangle + \alpha_2|2\rangle}{\sqrt{|\alpha_1|^2 + |\alpha_2|^2}}.$$

The measurement may also find the system in the subspace associated with Q with the probability $\langle\psi|Q|\psi\rangle$, which is $|\alpha_3|^2$, resulting in the post-measurement state $|e_3\rangle$.

2.4. **Observables.** An observable A is a measurable physical property which has a numerical value, for example, spin or momentum or energy. The term observable also extends to any basic measurement in which each outcome is associated with a numerical value. If λ_k is the numerical value associated with outcome $|e_k\rangle$ we say that observable A is represented by the operator

$$A = \sum_k \lambda_k |e_k\rangle\langle e_k| = \sum_k \lambda_k P_k,$$

where λ_k is now the eigenvalue corresponding to eigenvector $|e_k\rangle$, or projector P_k. Conversely, with any normal operator A we can associate a measurement defined by the eigenvectors of A, which form an orthonormal basis, and use the eigenvalues of A to label the outcomes of this measurement. If we choose the eigenvalues to be real numbers then A becomes a Hermitian operator. For example, the standard measurement on a single qubit is often called the "Z measurements" because the Pauli Z operator can be diagonalised in the standard basis and written as $Z = (+1)|0\rangle\langle 0| + (-1)|1\rangle\langle 1|$. The two outcomes, $|0\rangle$ and $|1\rangle$, are now labelled as $+1$ and -1, respectively. Using the same association we also have the X and the Y measurements, defined by the Pauli X and Y operators, respectively.

Operators

An operator A is said to be normal if $AA^\dagger = A^\dagger A$, unitary if $AA^\dagger = A^\dagger A = \mathbb{1}$, Hermitian or self-adjoint if $A^\dagger = A$, positive semidefinite if for any vector $|v\rangle$ we have $\langle v|A|v\rangle \geq 0$. A is normal if and only if it is unitarily diagonalisable. Both unitary and Hermitian operators are normal.

Let us mention in passing that many textbooks describe observables in terms of Hermitian operators and claim that the corresponding operators have to be Hermitian "because the outcomes are real numbers". This is misleading. The outcomes can be labelled by any symbols of your choice. It is the decomposition of the Hilbert space into mutually orthogonal subspaces that defines a measurement, not the labels. This said, labelling outcomes with real numbers is very useful. For example, the expectation value $\langle A\rangle$, which is the average of the numerical values λ_k weighted by their probabilities, is a very useful quantity and can be easily expressed in terms of the operator A as $\langle\psi|A|\psi\rangle$,

$$\begin{aligned}\sum_k \lambda_k \Pr(\text{outcome k}) &= \sum_k \lambda_k |\langle e_k|\psi\rangle|^2 = \sum_k \lambda_k \langle\psi|e_k\rangle\langle e_k|\psi\rangle \\ &= \langle\psi|\left(\sum_k \lambda_k |e_k\rangle\langle e_k|\right)|\psi\rangle = \langle\psi|A|\psi\rangle. \qquad (1)\end{aligned}$$

To be sure, this is not a value we expect to see in any particular experiment. Instead, imagine a huge number of quantum objects, all prepared in the state $|\psi\rangle$ and think about the observable A being measured on each of the objects. Statistically we then expect the average of our measurement results to be about $\langle A\rangle$. Note that when A is a projector then $\langle\psi|A|\psi\rangle$ is the probability of the outcomes associated with A.

2.5. **Compatible observables and the uncertainty relation.** (to be completed)

2.6. **Alice and Bob, and their quantum dramas.** (to be completed)

This is a good moment to introduce Alice and Bob (not their real names), our two protagonists who always need to communicate with each other. These two play the major role in many communication dramas, though they remain rather short on character development.

2.7. **Quantum communication.** This time Alice is sending quantum states to Bob and Bob does his best to identify them correctly by choosing appropriate measurements. Let us start we a simple observation, if a quantum state of the carrier of information is described by a state vector in a 2^n-dimensional Hilbert space then the carrier can carry at most n bits of information. For example, Alice can choose one of the 2^n states from a pre-agreed orthonormal basis $\{|e_k\rangle\}$ and Bob will be able to distinguish them reliably by choosing the $\{|e_k\rangle\}$ basis for his measurement. But can Alice and Bob do better than that? Can Alice send more than n bits of information per carrier by encoding them in states $|s_1\rangle, \ldots, |s_N\rangle$ where $N \geq 2^n$? Can Bob choose a clever measurement and reliably distinguish between all such states? The answer is no.

2.8. **Basic quantum coding and decoding.** Suppose Alice chooses randomly one of the pre-agreed N signal states $|s_k\rangle$ and sends it to Bob, whose tries to identify the signal states by performing a measurement defined by projectors P_l. Let P be a projector on a subspace that is spanned by the signal states $|e_k\rangle$, i.e. $P|s_k\rangle = |s_k\rangle$. The dimension of this subspace is given by $\mathrm{Tr}\, P = d$. We shall assume, without any loss of generality, that Bob designed his measurement in such a way that whenever he gets outcome P_k he concludes that Alice sent state $|s_k\rangle$. His probability of success is given by

$$\Pr(\text{success}) = \frac{1}{N}\sum_k \langle s_k| P_k |s_k\rangle ,$$

which is the probability that signal state $|s_k\rangle$ is selected, here $1/N$ for all the signal states are equally likely, times the probability that the selected signal state is correctly identified by Bob, which is $\langle s_k| P_k |s_k\rangle$, and we sum over all signal states.

Trace identities

$\mathrm{Tr}\,(ABC) = \mathrm{Tr}\,(CAB) = \mathrm{Tr}\,(BCA)$
$\mathrm{Tr}\,|a\rangle\langle b| = \langle b|a\rangle$
$\mathrm{Tr}\,|a\rangle\langle b|A = \langle b|A|a\rangle$
For any positive semidefinite B and projector P $\mathrm{Tr}\,BP \leq \mathrm{Tr}\,B$. To see this consider projector $Q = \mathbb{1} - P$ and notice that $\mathrm{Tr}\,B = \mathrm{Tr}\,B(P + Q) = \mathrm{Tr}\,BP + \mathrm{Tr}\,BQ$ and $\mathrm{Tr}\,BQ$ is non-negative.

Let us use this case to practice some of the trace identities, which we have listed in the margin. It is often convenient to write expressions such as $\langle\psi| A |\psi\rangle$ in terms of trace; for any vector $|\psi\rangle$ and operator A we have $\langle\psi| A |\psi\rangle = \mathrm{Tr}\,(A|\psi\rangle\langle\psi|) = \mathrm{Tr}\,(|\psi\rangle\langle\psi| A)$. In our case,

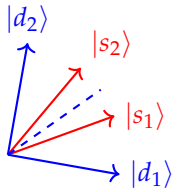
$$\Pr(\text{success}) = \frac{1}{N}\sum_k \langle s_k| P_k |s_k\rangle = \frac{1}{N}\sum_k \langle s_k| PP_kP |s_k\rangle = \frac{1}{N}\sum_k \mathrm{Tr}\,(PP_kP)\,|s_k\rangle\langle s_k| ,$$

where we have also used $P|s_k\rangle = |s_k\rangle$. Let us upper-bound this expression (see the box with trace identities),

$$\frac{1}{N}\sum_k \mathrm{Tr}\,(PP_kP)\,|s_k\rangle\langle s_k| \leq \frac{1}{N}\sum_k \mathrm{Tr}\,(PP_kP) = \frac{1}{N}\mathrm{Tr}\,(P(\sum_k P_k)P) = \frac{1}{N}\mathrm{Tr}\,(P\mathbb{1}P) = \frac{d}{N}.$$

Thus if Alice encodes N equally likely messages as states in a quantum system that, mathematically speaking, lives in the Hilbert space of dimension d, and if Bob decodes by performing a measurement and inferring the message from the result, Bob's probability of success is bounded by $\frac{d}{N}$. If the number N of possible signals exceeds the dimension d, Bob will not be able to reliably distinguish between the signals by any measurement. In particular, one qubit can store and carry at most one bit of information that can be reliably read by a measurement.

2.9. **Distinguishability of non-orthogonal states.** We have already mentioned that non-orthogonal states cannot be reliably distinguished, now it is time to make this statement more precise. Suppose Alices sends Bob a message by choosing one of the two non-orthogonal states, $|s_1\rangle$ or $|s_2\rangle$. The two messages are equally likely. What is the probability that Bob will decode the message correctly and what is the best measurement, i.e. the one that maximises this probability. As a general rule, before you embark on any calculations check for symmetries, special cases, anything that may help you to visualise the problem and make intelligent guesses about the solution. A good guess is one of the most powerful research tools. In fact, this is what real research is about, educated guesses that guide your calculations. In this particular case you can use symmetry arguments to guess the optimal measurement (see the margin). Once you guessed the answer, you might as well do the calculations. Suppose Bob's measurement is described by projectors P_1 and P_2, with the inference rule: P_1 implies $|s_1\rangle$ and P_2 implies $|s_2\rangle$.



The optimal measurement to distinguish between the two equally likely non-orthogonal signal states, $|s_1\rangle$ and $|s_2\rangle$, is described by the two orthogonal vectors, $|d_1\rangle$ and $|d_2\rangle$, placed symmetrically around the signal states.

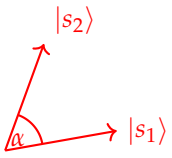
$$\Pr(\text{success}) = \frac{1}{2}(\langle s_1|\, P_1\, |s_1\rangle + \langle s_2|\, P_2\, |s_2\rangle) = \tfrac{1}{2}(\operatorname{Tr} P_1\, |s_1\rangle\langle s_1| + \operatorname{Tr} P_2\, |s_2\rangle\langle s_2|)$$
$$= \tfrac{1}{2}(\operatorname{Tr} P_1\, |s_1\rangle\langle s_1| + \operatorname{Tr}(\mathbb{1} - P_1)\, |s_2\rangle\langle s_2|) = \tfrac{1}{2}(1 + \operatorname{Tr} P_1(|s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|)). \quad (2)$$

Let us look at the operator $D = |s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|$ that appears in the last expression. This operator acts on the subspace spanned by $|s_1\rangle$ and $|s_2\rangle$, it is Hermitian and the sum of its two (real) eigenvalues is zero, $\operatorname{Tr} D = \langle\, s_1 \,|\, s_1 \,\rangle - \langle\, s_2 \,|\, s_2 \,\rangle = 0$. Let us express D as $\lambda(|d_+\rangle\langle d_+| - |d_-\rangle\langle d_-|)$, where $|d_\pm\rangle$ are the two orthonormal eigenstates of D and $\pm\lambda$ are the respective eigenvalues. Now we write

$$\Pr(\text{success}) = \tfrac{1}{2}(1 + \lambda \operatorname{Tr} P_1(|d_+\rangle\langle d_+| - |d_-\rangle\langle d_-|)) \le \tfrac{1}{2}(1 + \lambda\, \langle d_+|\, P_1\, |d_+\rangle), \quad (3)$$

where we have dropped the non-negative term $\operatorname{Tr} P_1\, |d_-\rangle\langle d_-|$. In fact, it is easy to see that we will maximise the expression above by choosing $P_1 = |d_+\rangle\langle d_+|$ (and $P_2 = |d_-\rangle\langle d_-|$). The probability of success is then bounded by $\frac{1}{2}(1+\lambda)$. All we have to do now is to find the positive eigenvalue λ for the operator D. You can do it, of course, by solving the characteristic equation for a matrix representation of D but, as we are now practicing the trace operations, we can also notice that $\operatorname{Tr} D^2 = 2\lambda^2$ and evaluate the trace of D^2. We use the trace identities and obtain $\operatorname{Tr} D^2 = \operatorname{Tr}(|s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|)(|s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|) = 2 - 2|\langle\, s_1 \,|\, s_2 \,\rangle|^2$, which gives $\lambda = \sqrt{1 - |\langle\, s_1 \,|\, s_2 \,\rangle|^2}$. Bringing it all together we have the final expression

$$\Pr(\text{success}) = \tfrac{1}{2}(1 + \sqrt{1 - |\langle\, s_1 \,|\, s_2 \,\rangle|^2}).$$



We can parametrise $|\langle\, s_1 \,|\, s_2 \,\rangle| = \cos\alpha$ and interpret α as the angle between $|s_1\rangle$ and $|s_2\rangle$. This allows us to express our findings in a more clear way. Given two equally likely states $|s_1\rangle$ and $|s_2\rangle$, such that $|\langle\, s_1 \,|\, s_2 \,\rangle| = \cos\alpha$, the probability of correctly identifying the state by a projective measurement is bounded by

$$\Pr(\text{success}) = \tfrac{1}{2}(1 + \sin\alpha),$$

and the optimal measurement, that achieves this bound, is determined by the eigenvectors of $D = |s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|$ (try to visualise these eigenvectors). It makes sense, right? If we try just guessing the state, without any measurement, we expect $\Pr(\text{success}) = \frac{1}{2}$. This is our lower bound and in any attempt to distinguish the two states we should do better than that. If the two signal states are very close to each other than $\sin\alpha$ is small and we are slightly better off than guessing. As we increase α the two states become more distinguishable, and, as we can see from the formula, when the two states become orthogonal they also become completely distinguishable.

2.10. **Wiesner's quantum money.** (to be completed)

Notes & Exercises

(1) Consider two unit vectors $|a\rangle$ and $|b\rangle$. Is $|a\rangle\langle a| + |b\rangle\langle b|$ a projector?

(2) Suppose you are given a single qubit in some unknown quantum state $|\psi\rangle$. Can you determine $|\psi\rangle$?

(3) You measure a random qubit in the standard basis and register $|0\rangle$. What does it tell you about the pre-measurement state $|\psi\rangle$?

(4) How many real parameters do you need to determine $|\psi\rangle$? Would you be able to reconstruct $|\psi\rangle$ from $\langle\psi|X|\psi\rangle$, $\langle\psi|Y|\psi\rangle$ and $\langle\psi|Z|\psi\rangle$? It may help you to visualise $|\psi\rangle$ as a Bloch vector.

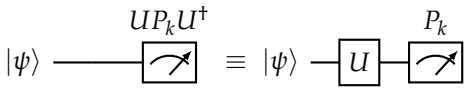
(5) You are given zillions of qubits, all prepared in the same quantum state $|\psi\rangle$. How would you determine $|\psi\rangle$

(6) The Z measurement is defined by the projectors

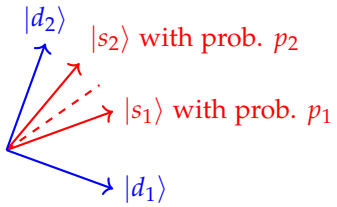
$$P_0 = \tfrac{1}{2}(\mathbb{1} + Z) \qquad P_1 = \tfrac{1}{2}(\mathbb{1} - Z).$$

Consider a measurement associated with Hermitian operator S, such that $S^2 = \mathbb{1}$. Show that the two outcomes ± 1 correspond to the projectors $\frac{1}{2}(\mathbb{1} \pm S)$.

(7) In our quantum circuits, unless specified otherwise, all measurements are assumed to be performed in the standard basis. This is because any measurement can be reduced to the standard measurement by performing some prior unitary transformation. Show that any two orthonormal bases $\{|e_k\rangle\}$ and $\{|d_l\rangle\}$ are always related by some unitary U (i.e. show that $\sum_k |d_k\rangle\langle e_k|$ is unitary). Suppose projectors P_k define the standard measurement, show that for any unitary U projectors $UP_kU^\dagger$ also define a measurement.



(8) The optimal measurement to distinguish between the two equally likely non-orthogonal signal states, $|s_1\rangle$ and $|s_2\rangle$, is described by the two orthogonal vectors, $|d_1\rangle$ and $|d_2\rangle$, placed symmetrically around the signal states. But suppose the states are not equally likely; $|s_1\rangle$ is chosen with probability p_1 and $|s_2\rangle$ with probability p_2. How would you modify the measurement to maximise the probability of success in this case?



(9) **How to ascertain the values of σ_x and σ_y of a qubit?** Alice prepares a qubit in any state of her choosing and gives it to Bob who secretly measures either σ_x or σ_y. The outcome of the measurement is seen only by Bob. Alice has no clue which measurement was chosen by Bob but right after his measurement she gets her qubit back and she can measure it as well. Some time later Bob tells Alice which of the two measurements was chosen, i.e. whether he measured σ_x or σ_y. Alice then tells him the outcome he obtained in his measurement. Bob is surprised for the two measurements have mutually unbiased bases and yet Alice always gets it right. How does she do it?

This is a simplified version of a beautiful quantum puzzle proposed in 1987 by Lev Vaidman, Yakir Aharonov, and David Z. Albert in a paper with the somewhat provocative title, "How to ascertain the values of σ_x, σ_y, and σ_z of a spin-$\frac{1}{2}$ particle." For the original see Phys. Rev. Lett. vol. 58, 1385 (1987).

(10) **Zeno effect** (to be completed)

Quantum theory revisited

Even though multiplying and adding probability amplitudes is essentially all there is to quantum theory we hardly ever multiply and add amplitudes in a pedestrian way. Instead, as we have seen, we neatly tabulate the amplitudes into vectors and matrices and let the matrix multiplication take care of multiplication and addition of amplitudes corresponding to different alternatives. Thus vectors and matrices appear naturally as our bookkeeping tools; we use vectors to describe quantum states and matrices (operators) to describe quantum evolutions and measurements. This leads to a convenient mathematical setting for quantum theory, which is a complex vector space with an inner product, often referred to as a Hilbert space. It turns out, somewhat miraculously, that this pure mathematical construct is exactly what we need to formalise quantum theory. It gives us a precise language which is appropriate for making empirically testable predictions. At a very instrumental level, quantum theory is a set of rules designed to answer questions such as 'given a specific preparation and a subsequent evolution compute probabilities for the outcomes of such and such measurement'. Here is how we represent preparations, evolutions and measurements in mathematical terms, and how we get probabilities.

2.11. **Quantum states.** With any isolated quantum system, which can be prepared in n perfectly distinguishable states, we can associate a Hilbert space $\mathcal{H}$ of dimension n such that each vector $|v\rangle \in \mathcal{H}$ of unit length, $\langle v \,|\, v \rangle = 1$, represents a quantum state of the system. The overall phase of the vector has no physical significance: $|v\rangle$ and $e^{i\varphi}|v\rangle$, for any real φ, describe the same state. The inner product $\langle u \,|\, v \rangle$ is the probability amplitude that a quantum system prepared in state $|v\rangle$ will be found in state $|u\rangle$. States corresponding to orthogonal vectors, $\langle u \,|\, v \rangle = 0$, are perfectly distinguishable for the system prepared in state $|v\rangle$ will never be found in state $|u\rangle$, and vice versa. In particular, states forming orthonormal bases are always perfectly distinguishable from each other.

2.12. **Quantum evolutions.** Any physically admissible evolution of an isolated quantum system is represented by a unitary operator. Unitary operators describing evolutions of quantum systems are usually derived from the Schrödinger equation,

$$\frac{d}{dt}\,|\psi(t)\rangle = -\frac{i}{\hbar}\,H\,|\psi(t)\rangle\,, \tag{4}$$

Here $\hbar = 1.05 \times 10^{-34}$ J s denotes Planck's constant. Theorists will always choose to work with a system of units where $\hbar = 1$.

where H is a Hermitian operator called the Hamiltonian. It contains a complete specification of all interactions both within the system and between the system and the external potentials. For time independent Hamiltonians the formal solution of the Schrödinger equation reads

$$|\psi(t)\rangle = U(t)\,|\psi(0)\rangle \qquad \text{where} \qquad U(t) = e^{-\frac{i}{\hbar}Ht} \tag{5}$$

Any unitary matrix can be represented as the exponential of some Hermitian matrix, H and a real coefficient t,

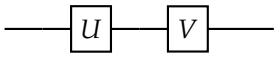
We shall ignore the convergence issues

$$e^{itH} \equiv \mathbb{1} + itH + \frac{(it)^2}{2}H^2 + \frac{(it)^3}{2\cdot 3}H^3 ... = \sum_{n=0}^{\infty}\frac{(it)^n}{n!}\,H^n. \tag{6}$$

The state vector changes smoothly; for $t = 0$ the time evolution operator is merely the unit operator $\mathbb{1}$, and when t is very small $U(t) \approx \mathbb{1} - itH$ is close to the unit operator, differing from it by something of order t.

2.13. **Quantum circuits.** In this course we will hardly refer to the Schrödinger equation, instead we will assume that our clever colleagues, experimental physicists, are able to implement certain unitary operations and we will use these unitaries, like lego blocks, to construct other, more complex, unitaries. We refer to preselected

elementary quantum operations as quantum logic gates and we often draw diagrams, called quantum circuits, to illustrate how they act on qubits. For example, two unitaries, U followed by V, acting on a single qubit are represented as



This diagram should be read from left to right. The horizontal line represents a qubit that is inertly carried from one quantum operation to another.

2.14. **Measurements.** A complete measurement in quantum theory is determined by the choice of an orthonormal basis $\{|e_i\rangle\}$ in $\mathcal{H}$, and every such basis in principle represents a possible measurement. Given a quantum system in state $|\psi\rangle$, such that

$$|\psi\rangle = \sum_i |e_i\rangle \langle e_i \,|\, \psi \rangle$$

the measurement in the basis $\{|e_i\rangle\}$ gives outcome labelled by e_k with probability $|\langle e_k \,|\, \psi \rangle|^2$ and leaves the system in state $|e_k\rangle$. This is consistent with our interpretation of the inner product $\langle e_k \,|\, \psi \rangle$ as the probability amplitude that a quantum system prepared in state $|\psi\rangle$ will be found in state $|e_k\rangle$. State vectors forming orthonormal bases are perfectly distinguishable from each other, $\langle e_i \,|\, e_j \rangle = \delta_{ij}$, hence there is no ambiguity about the outcome. A complete measurement is the best we can do in terms of resolving state vectors in the basis states. In general, for any decomposition of the identity $\sum_k P_k = \mathbb{1}$ into orthogonal projectors P_k ($P_k P_l = P_k \delta_{kl}$) there exists a measurement that takes a quantum system in state $|\psi\rangle$, outputs label k, with probability $\langle\psi| P_k |\psi\rangle$ and leaves the system in the state $P_k |\psi\rangle$ (multiplied by the normalisation factor i.e. divided by the length of $P_k |\psi\rangle$),

$$|\psi\rangle \longrightarrow \frac{P_k |\psi\rangle}{\sqrt{\langle\psi| P_k |\psi\rangle}}.$$

The projector formalism covers both complete and incomplete measurements. The complete measurements are defined by rank one projectors, $P_k = |e_k\rangle \langle e_k|$, projecting on vectors from some orthonormal basis $|e_k\rangle$.

BELL'S THEOREM

ARTUR EKERT

Introduction to Quantum Information Science
Lecture 4

About quantum correlations, which are stronger than any correlations allowed by classical physics, and about the CHSH inequality which demonstrates this fact.

1.1. **Quantum correlations.** Consider two entangled qubits in the singlet state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right),$$

and note that the projector $|\psi\rangle\langle\psi|$ can be written as

$$|\psi\rangle\langle\psi| = \frac{1}{4}\left(\mathbb{1}\otimes\mathbb{1} - \sigma_x\otimes\sigma_x - \sigma_y\otimes\sigma_y - \sigma_z\otimes\sigma_z\right).$$

There are other, more elementary, ways of deriving this result but here I want you to hone your skills. Once you learned about projectors, traces and Pauli operators why not putting them into good use.

Any single qubit observable with values ± 1 can be represented by the operator $\vec{a}\cdot\vec{\sigma}$,

$$\vec{a}\cdot\vec{\sigma} = a_x\sigma_x + a_y\sigma_y + a_z\sigma_z,$$

where $\vec{a}$ is a unit vector in the three-dimensional Euclidean space. Suppose Alice and Bob choose measurements defined by vectors $\vec{a}$ and $\vec{b}$, respectively. For example, if the two qubits are spin-half particles they may measure the spin components along the directions $\vec{a}$ and $\vec{b}$. We write the corresponding observable as the tensor product,

$$A\otimes B = (\vec{a}\cdot\vec{\sigma})\otimes(\vec{b}\cdot\vec{\sigma}).$$

The eigenvalues of $A\otimes B$ are the products of eigenvalues of A and B. Thus $A\otimes B$ has two eigenvalues, $+1$ corresponding to the instances when Alice and Bob registered identical outcomes $(+1,+1)$ or $(-1,-1)$, and -1 corresponding to the instances when Alice and Bob registered different outcomes$(+1,-1)$ or $(-1,-1)$. This means that the expectation value of $A\otimes B$, in any state, has a simple interpretation,

$$\langle A\otimes B\rangle = \Pr(\text{outcomes are the same}) - \Pr(\text{outcomes are different}).$$

This expression can take any numerical value from -1 (perfect anti-correlations) through 0 (no correlations) to $+1$ (perfect correlations). We now evaluate the expectation value in the singlet state

$$\begin{aligned}
\langle\psi|\, A\otimes B\,|\psi\rangle &= \mathrm{Tr}\left[(\vec{a}\cdot\vec{\sigma})\otimes(\vec{b}\cdot\vec{\sigma})\ \ |\psi\rangle\langle\psi|\right] \\
&= -\frac{1}{4}\mathrm{Tr}\left[(\vec{a}\cdot\vec{\sigma})\sigma_x\otimes(\vec{a}\cdot\vec{\sigma})\sigma_x + (\vec{a}\cdot\vec{\sigma})\sigma_y\otimes(\vec{a}\cdot\vec{\sigma})\sigma_y + (\vec{a}\cdot\vec{\sigma})\sigma_z\otimes(\vec{a}\cdot\vec{\sigma})\sigma_z\right] \\
&= -\frac{1}{4}\mathrm{Tr}\left[(a_xb_x + a_yb_y + a_zb_z)\,\mathbb{1}\otimes\mathbb{1}\right] \\
&= -\vec{a}\cdot\vec{b} \qquad (1)
\end{aligned}$$

where we have used $\mathrm{Tr}\,(\vec{a}\cdot\vec{\sigma})\sigma_k = a_k$ $(k = x,y,z)$. Thus, if Alice and Bob choose the same observable, $\vec{a} = \vec{b}$, their outcomes will be always opposite; whenever Alice registers $+1$ (-1) Bob is bound to register -1 $(+1)$.

1.2. **Hidden variables.** The story of "hidden variables" dates back to 1935 and grew out of Einstein's worries about the completeness of quantum theory. Consider, for example, a qubit. No quantum state of a qubit can be an eigenstate of two non-commuting operators, say σ_x and σ_z. If the qubit has a definite value of σ_x its value of σ_z must be indeterminate, and vice versa. If we take quantum theory to be a complete description of the world, then we must accept that it is impossible for both σ_x and σ_z to have definite values for the same qubit at the same time. Einstein felt very uncomfortable about all this. He argued that quantum theory is incomplete, that observables σ_x and σ_z both may have simultaneous definite values, although we only have knowledge of one of them at a time. This is the hypothesis of "hidden variables". In this view, the indeterminacy found in quantum theory is merely due to our ignorance of these "hidden variables" that are present in nature but not accounted for in the theory. Einstein came up with a number of pretty good arguments for the existence of "hidden variables". Probably the most compelling one was described in his 1935 paper, co-authored with his younger colleagues, Boris Podolsky and Nathan Rosen (known as the EPR paper). It stood for almost three decades as the most significant challenge to the completeness of quantum theory. Then, in 1964, John Bell showed that the hidden variable hypothesis can be tested and refuted.

UPPER BOUND ON CLASSICAL CORRELATIONS

1.3. **CHSH inequality.** I will describe the most popular version of Bell's argument, introduced in 1969 by John Clauser, Michael Horne, Abner Shimony and Richard Holt (CHSH). Let us assume that the results of any measurement on any individual system are predetermined. Any probabilities we may use to describe the system merely reflect our ignorance of these hidden variables.

Now, imagine the following scenario. Alice and Bob, two characters with a predilection for wacky experiments, are equipped with appropriate measuring devices and sent to two distant locations. Somewhere in between them there is a source that emits pairs of qubits that fly apart, one towards Alice and one towards Bob. Let us label the two qubits in each pair as $\mathcal{A}$ and $\mathcal{B}$ respectively and let us assume that both A and B have well defined values of their observables. We ask Alice and Bob to measure one of the two pre-agreed observables. For each incoming qubit, Alice and Bob choose randomly, and independently from each other, which particular observable will be measured. Alice chooses between A_1 and A_2, and Bob between B_1 and B_2. Each observable has value $+1$ or -1 thus we are allowed to think about them as random variables A_k and B_k, $k = 1, 2$, which take values ± 1. Let us define a new random variable, the CHSH quantity S,

$$S = A_1(B_1 - B_2) + A_2(B_1 + B_2).$$

It is easy to see that one of the terms $B_1 \pm B_2$ must be equal to zero and the other to ± 2, hence $S = \pm 2$. The average value of S must lie somewhere in-between, i.e.

$$-2 \leq \langle S \rangle \leq 2. \tag{2}$$

That's it! Such a simple and yet profound mathematical statement about correlations, to which we refer simply as the CHSH inequality. No quantum theory involved because the CHSH inequality is not specific to quantum theory; it does not really matter what kind of physical process is behind the appearance of binary values of A_1, A_2, B_1 and B_2. It is a statement about correlations and for all classical correlations we must have

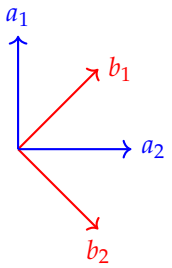
$$|\langle A_1 B_1 \rangle - \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle + \langle A_2 B_2 \rangle| \leq 2.$$

There are essentially two two assumptions here:

(1) **Hidden variables** Observables have definite values
(2) **Locality** Alice's choice of measurements (A_1 or A_2) does not affect the outcomes of Bob's measurement, and vice versa.

I will not discuss the locality assumption here in detail but let me comment on it briefly. In the hidden variable world a statement such as "if Bob were to measure B_1 then he would register $+1$" must be either true or false prior to Bob's measurement. Without the locality hypothesis such a statement is ambiguous, since the value of B_1 could depend on whether A_1 or A_2 will be chosen by Alice. We do not want this for it implies the instantaneous communication. It means that, say, Alice by making a choice between A_1 and A_2, affects Bob's results. Bob can immediately 'see' what Alice 'does'.

1.4. **Quantum correlations revisited.** In quantum theory the observables A_1, A_2, B_1, B_2 become 2×2 Hermitian matrices with two eigenvalues ± 1, and $\langle S \rangle$ becomes the expectation value of the 4×4 CHSH matrix



$$S = A_1 \otimes (B_1 - B_2) + A_2 \otimes (B_1 + B_2).$$

We can now evaluate $\langle S \rangle$ using quantum theory. For example, if the two qubits are in the singlet state we know that $\langle A \otimes B \rangle = -\vec{a} \cdot \vec{b}$. We choose vectors $\vec{a}_1$, $\vec{a}_2$, $\vec{b}_1$ and $\vec{b}_2$ as shown in the picture (the relative angle between the two perpendicular pairs is 45 degrees) and with these choices

$$\langle A_1 \otimes B_1 \rangle = \langle A_2 \otimes B_1 \rangle = \langle A_2 \otimes B_2 \rangle = \frac{1}{\sqrt{2}}, \qquad \langle A_1 \otimes B_2 \rangle = -\frac{1}{\sqrt{2}}.$$

Thus

$$\langle A_1 B_1 \rangle - \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle + \langle A_2 B_2 \rangle = -2\sqrt{2},$$

which obviously violates CHSH inequality. And this, to be sure, has been observed in a number of painstaking experiments. The early efforts were truly heroic, and the experiments had many layers of complexity. Today, however, such experiments are routine. The behaviour of entangled quantum systems cannot be explained by any local hidden variables.

UPPER BOUND ON QUANTUM CORRELATIONS

1.5. **Tsirelson's inequality.** One may ask, and indeed one of you asked, if $|\langle S \rangle| = 2\sqrt{2}$ is the maximal violation of the CHSH inequality. Yes, it is. Quantum correlations cannot achieve any larger value of $|\langle S \rangle|$ than $2\sqrt{2}$. This is because for any state $|\psi\rangle$ the expectation value $\langle S \rangle = \langle \psi | S | \psi \rangle$ cannot exceed the largest eigenvalue of S and we can put an upper bound on the largest eigenvalues of S. To start with, the largest eigenvalue (in absolute value) of a Hermitian matrix M, denoted by $||M||$, is a matrix norm, and it has the following properties:

$$||M \otimes N|| = ||M||\,||N||, \quad ||MN|| \le ||M||\,||N||, \quad ||M + N|| \le ||M|| + ||N||.$$

Given that $||A_i|| = 1$ and $||B_j|| = 1$ $(i, j = 1, 2)$ it is easy to show that $||S|| \le 4$. One can, however, derive a tighter bound. We can show (do it) that

$$S^2 = 4\, \mathbb{1} \otimes \mathbb{1} + [A_1, A_2] \otimes [B_1, B_2].$$

The norm of each of the commutators, $||[A_1, A_2]||$ and $||[B_1, B_2]||$, cannot exceed 2 and $||S^2|| = ||S||^2$, which all together gives

$$||S|| \le 2\sqrt{2}, \quad \text{which implies} \quad |\langle S \rangle| \le 2\sqrt{2}.$$

This result is known as the Tsirelson inequality.

QUANTUM ALGORITHMS

ARTUR EKERT

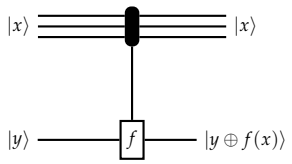
It is time to start talking about quantum computation per se...

2.1. **Quantum Boolean function evaluation.** Classical computers essentially evaluate functions: given n-bits of input they produce m-bits of output that are uniquely determined by the input; that is, they find the value of

$$f : \{0,1\}^n \to \{0,1\}^m$$

for a particular specified n-bit argument. A function with an m-bit value is equivalent to m Boolean functions, each with a one-bit value, so we may just as well say that the basic task performed by a computer is the evaluation of Boolean functions,

$$f : \{0,1\}^n \to \{0,1\}.$$



In quantum computation all elementary operations are reversible (unitary) so we compute Boolean functions in a reversible fashion as

$$|x\rangle\,|y\rangle \mapsto |x\rangle\,|y \oplus f(x)\rangle\,.$$

The corresponding circuit diagram (for $n = 3$) is shown in the margin. Here we use two registers, the first one (counting from the top to the bottom of the circuit diagram) stores the arguments x and the second one the values $f(x)$. More precisely, the value $f(x)$ is added bit-wise to a pre-existing binary value y of the second register. We usually set $y = 0$ to get

$$|x\rangle\,|0\rangle \mapsto |x\rangle\,|f(x)\rangle\,.$$

Quantum Boolean function evaluation is a special case of the generalised x-controlled-U on two registers,

$$\sum_x |x\rangle\,\langle x| \otimes U_x,$$

where U_x is either the identity $\mathbb{1}$ (when $f(x) = 0$) or the bit-flip X (when $f(x) = 1$). (Please, do not confuse here the capital X, which is the Pauli flip operator σ_x, with the small x, which is a binary string stored in the first register and the argument of our Boolean function f.) We can also write it as

$$\sum_x |x\rangle\,\langle x| \otimes X^{f(x)}.$$

2.2. **Example.** Consider a Boolean function $f : \{0,1\}^2 \to \{0,1\}$ such that $f(x) = 1$ for $x = 01$ and $f(x) = 0$ otherwise. The evaluation $|x\rangle\,|y\rangle \mapsto |x\rangle\,|y \oplus f(x)\rangle$ can be written explicitly as

x	$f(x)$
00	0
01	1
10	0
11	0

$$\begin{array}{lcllcl}
|00\rangle\,|0\rangle & \mapsto & |00\rangle\,|0\rangle & |00\rangle\,|1\rangle & \mapsto & |00\rangle\,|1\rangle \\
|01\rangle\,|0\rangle & \mapsto & |01\rangle\,|1\rangle & |01\rangle\,|1\rangle & \mapsto & |01\rangle\,|0\rangle \\
|10\rangle\,|0\rangle & \mapsto & |10\rangle\,|0\rangle & |10\rangle\,|1\rangle & \mapsto & |10\rangle\,|1\rangle \\
|11\rangle\,|0\rangle & \mapsto & |11\rangle\,|0\rangle & |11\rangle\,|1\rangle & \mapsto & |11\rangle\,|1\rangle
\end{array}$$

and the expression $\sum_x |x\rangle\,\langle x| \otimes X^{f(x)}$ becomes

$$|00\rangle\,\langle 00|\,00 \otimes \mathbb{1} + |01\rangle\,\langle 01|\,01 \otimes X + |10\rangle\,\langle 10|\,10 \otimes \mathbb{1} + |11\rangle\,\langle 11|\,11 \otimes \mathbb{1}.$$

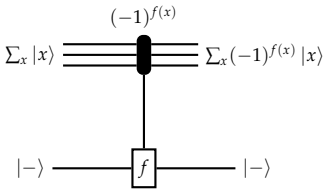
Finally, in the matrix form,

$$\left[\begin{array}{cc|cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

As you can see, this is a diagonal block matrix; the 4×4 matrix with 2×2 matrices as entries. The rows and the columns of the 4×4 matrix are labelled by the binary strings $00, 01, 10, 11$, and the 2×2 matrices on the diagonal represent operations applied to the qubit in the second register. Here all of them are the identity $\mathbb{1}$ except the $(01, 01)$ entry which represents bit-flip X; this is because $f(01) = 1$ and $f(x) = 0$ for all other binary strings x.

2.3. **Phase kick-back.** What makes the quantum evaluation of Boolean functions really interesting is its action on a superposition of different inputs x. For example,

$$\sum_x |x\rangle\,|0\rangle \mapsto \sum_x |x\rangle\,|f(x)\rangle$$



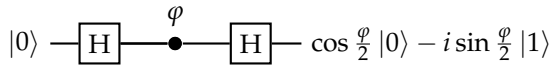
produces $f(x)$ for all x in a single run (we have dropped the normalisation factor). It is more instructive to see the effect of the function evaluation when the qubit in the second register is prepared in the state $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$,

$$\sum_x |x\rangle\,|-\rangle \mapsto \sum_x (-1)^{f(x)}\,|x\rangle\,|-\rangle\,.$$

Whenever $f(x) = 1$ the bit flip X is applied to the qubit in the second register. The state $|-\rangle$ is the eigenstate of X with eigenvalue -1, thus, due the phase kick-back, whenever $f(x) = 1$ the phase factor -1 appears in front of the corresponding term $|x\rangle$. As you can see, the second register stays in state $|-\rangle$ all way through the computation. It is the first register where things happen. Let us see now how quantum Boolean function evaluation introduces phase shifts in quantum interference experiments and how such experiments can be viewed as computations.

2.4. **Oracles and query complexity.** The computational power of quantum interference was discovered by counting how many times certain Boolean functions have to be evaluated in order to find the answer to a given problem. Imagine a "black box" (also called an *oracle*) computing a Boolean function and a scenario in which one wants to learn about a given property of the Boolean function but has to pay for each use of the box (often referred to as a *query*). In such a setting, the objective is to minimise number of queries to the oracle. We are ignoring everything that happens inside the black box — the Boolean function evaluation is just one computational step.

2.5. **Deutsch's algorithm.** We start with the simplest quantum interference circuit,

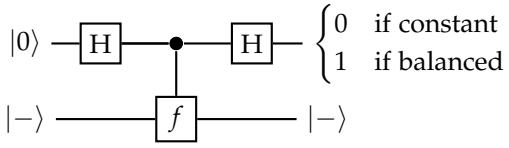


Suppose you can prepare the input and read the output but you cannot see the phase shifter, but you are promised that the phase shifter is set either to $\phi = 0$ or to $\phi = \pi$. Can you tell the difference? Of course you can. One way of doing it is to set your input to $|0\rangle$ and check the output; for $\phi = 0$ the output is aways $|0\rangle$ and for $\phi = \pi$ it is always $|1\rangle$. A single run of the interference experiment is sufficient to determine the difference. The first quantum algorithm, proposed by David Deutsch

in 1985, is very much related to this effect. The phase setting is determined by the Boolean function evaluation via the phase kick-back.

Consider the Boolean functions f that map $\{0,1\}$ to $\{0,1\}$. There are exactly four such functions: two constant functions ($f(0) = f(1)$) and two "balanced" functions ($f(0) \neq f(1)$). Informally, in Deutsch's problem, one is allowed to evaluate the function f *only once* and required to deduce from the result whether f is constant or balanced. Note that we are not asked for the particular values $f(0)$ and $f(1)$ but only whether the two values are the same or different. Classical intuition tells us that we have to evaluate both $f(0)$ and $f(1)$ and compare them, which involves evaluating f twice. In the quantum setting we can solve this problem with a single function evaluation, as illustrated by the circuit below.

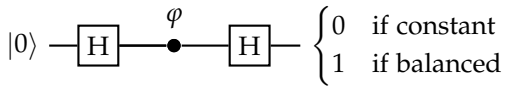
$f(0)$	$f(1)$	
0	0	contant
1	1	contant
0	1	balanced
1	0	balanced



During the function evaluation the second register "kicks back" the phase factor $(-1)^{f(x)}$ in front of $|x\rangle$. The state of the second register remains unchanged while the first is modified as follows

$$\begin{aligned} |0\rangle &\overset{H}{\mapsto} |0\rangle + |1\rangle \\ &\overset{f}{\mapsto} (-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle \\ &\equiv |0\rangle + (-1)^{f(0)\oplus f(1)}|1\rangle \\ &\overset{H}{\mapsto} |f(0)\oplus f(1)\rangle \end{aligned}$$

This evolution can be represented by the circuit diagram,



where the relative phase $\varphi = (-1)^{f(0)\oplus f(1)}$. The first qubit ends in state $|0\rangle$ if the function f is constant and in state $|1\rangle$ if the function is balanced, and the standard measurement distinguishes these two cases with certainty.

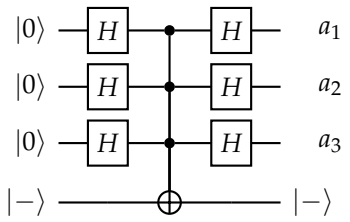
Deutsch's result laid the foundation for the new field of quantum computation, and was followed by several other quantum algorithms for various problems. They all seem to rest on the same generic sequence: a Hadamard transform, followed by a function evaluation, followed by another Hadamard (or Fourier) transform. As we shall see in a moment, in some cases, such as Grover's search algorithm, this sequence is repeated several times. Let me now take you through the three early quantum algorithms.

The Hadamard transform is a special case of the Fourier transform; it is the Fourier transform over group $\mathbb{Z}_2^n$.

2.6. **Bernstein-Vazirani algorithm.** We are presented with an oracle, i.e. a black box whose internal processes we don't care about, which evaluates $f : \{0,1\}^n \to \{0,1\}$, and we are promised that f is of the form

$$f(x) = a \cdot x \equiv (a_1 \cdot x_1) \oplus \cdots \oplus (a_n \cdot x_n)$$

where $a \in \{0,1\}^n$. Our task is to determine the value of the n-bit string a while minimising the number of calls to the oracle. It's quite easy to see how to do this classically; if we input a value $x = 00\ldots010\ldots0$, with the 1 on bit m, then $f(x)$ is simply the m^{th} bit of a. After n similar calls, we can evaluate every bit value. It is also clear that there cannot exist a better classical algorithm – each call to the oracle teaches us exactly one bit of information, and since we must learn n bits, we must query it n times. In contrast, by running the quantum circuit shown below it is possible to determine a with a single call to the oracle.



Stepping through the execution of the circuit we obtain (we neglect the second register that remains in state $|-\rangle$ all way through)

$$
\begin{aligned}
|0\rangle &\overset{H}{\mapsto} \left(\frac{1}{\sqrt{2}}\right)^n \sum_{x\in\{0,1\}^n} |x\rangle \\
&\overset{f}{\mapsto} \left(\frac{1}{\sqrt{2}}\right)^n \sum_{x\in\{0,1\}^n} (-1)^{a\cdot x} |x\rangle \\
&\overset{H}{\mapsto} \left(\frac{1}{\sqrt{2}}\right)^n \sum_{x\in\{0,1\}^n} (-1)^{a\cdot x} \left[\left(\frac{1}{\sqrt{2}}\right)^n \sum_{y\in\{0,1\}^n} (-1)^{y\cdot x} |y\rangle\right] \\
&= \left(\frac{1}{2}\right)^n \sum_{y\in\{0,1\}^n} \left[\sum_{x\in\{0,1\}^n} (-1)^{(a\oplus y)\cdot x}\right] |y\rangle = |a\rangle\,,
\end{aligned}
$$

where the second Hadamard transform is written as

$$
|x\rangle \mapsto \left(\frac{1}{\sqrt{2}}\right)^n \sum_{y\in\{0,1\}^n} (-1)^{y\cdot x} |y\rangle
$$

and we have used the identity

$$
\sum_{x\in\{0,1\}^n} (-1)^{x\cdot y} = \begin{cases} 0 & \text{if } y \neq 0 \\ 2^n & \text{if } y = 0 \end{cases}
$$

for any $y \in \{0,1\}^n$. This identity (prove it) allowed us to write

$$
\sum_{x\in\{0,1\}^n} (-1)^{(a\oplus y)\cdot x} = \begin{cases} 0 & \text{if } y \neq a \\ 2^n & \text{if } y = a \end{cases}
$$

Even if you don't instantly see how this sum works for $z \neq a$ you can first calculate the probability that the output is $z = a$. In this case it is easy to see that the sum is 2^n and that in the final state $\sum_z \alpha_z |z\rangle$ the term $z = a$ has amplitude 1. Thus, due to the normalisation, all the others terms must be equal to 0.

If you take the sum over x, then all the terms always cancel out unless $a \oplus y = 00\ldots0$ i.e. $y = a$. Thus the standard bit by bit measurement of the first register gives the value of a and solves the problem with a single call to the oracle.

3. Simon's problem

Simon's Problem is the simplest quantum algorithm that shows an exponential oracle-based speed-up over the best classical algorithm. Suppose we are given an oracle, a black box that computes unknown function $f : \{0,1\}^n \to \{0,1\}^n$, and we are assured that f is two-to-one: $f(x \oplus s) = f(x)$ for some fixed binary string s. For simplicity, in our first approach, we shall assume that s is different from zero ($s \neq 0^n$). Our task is to find s, again with as few calls to the oracle as possible. For example, for $n = 3$ the oracle may compute the function shown in the table (in the margin). Every output of f occurs twice, and the bitwise sum of two input strings corresponding to any one given output equals to $s = 101$.

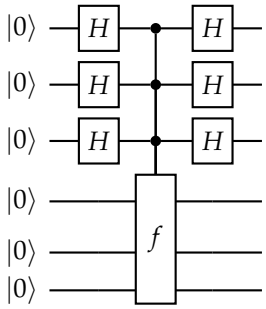
x	$f(x)$
000	110
001	111
010	001
011	000
100	111
101	110
110	000
111	001

$f(x) = f(x \oplus s)$ for $s = 101$

Classically, this problem is exponentially hard. We will not go through a detailed proof of this fact, but the intuition is reasonably simple: since there is not any structure in the function f that would help you to find its period s, the best we can do is to evaluate f on random inputs and if we are ever lucky enough to find x and y such that $f(x) = f(y)$, we get our answer $s = x \oplus y$. After having made m queries to the oracle, we have a list of m values of x and $f(x)$. There are $\frac{1}{2}m(m-1)$ possible

pairs which could match within this set, and the probability that a randomly chosen pair match is $1/2^{n-1}$. This means that the probability of there being at least one matching pair within the set of m strings is less than $m^2/2^n$. Clearly, the chance of finding a matching pair is negligible if f is probed on fewer than $\sqrt{2^n}$ inputs.

The quantum case, on the other hand, gives a result with high probability within a linear number of steps. The circuit that solves the problem has a familiar Hadamard-function-Hadamard structure but the second register has now expanded to n qubits.



We prepare the first register in the equally weighted superposition of all n-bit strings,

$$|0^n\rangle\,|0^n\rangle \to \frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|x\rangle\,|0^n\rangle\,,$$

and query the oracle,

$$\to \frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|x\rangle\,|f(x)\rangle\,.$$

The second Hadamard transform on the first register yields the final output state

$$\frac{1}{2^n}\sum_{x,y}(-1)^{x\cdot y}\,|y\rangle\,|f(x)\rangle\,. \tag{1}$$

If we measure the second register we obtain one of the 2^{n-1} possible values of $f(x)$, each equally likely. Suppose the outcome is $f(a)$. Given that both a and $a\oplus s$ are mapped by f to $f(a)$ the first registered ends up in the state

As we shall see in a moment, the actual measurement on the second register is not necessary.

$$\frac{1}{\sqrt{2}}(|a\rangle + |a\oplus s\rangle).$$

The subsequent Hadamard transform on the first register gives

$$\frac{1}{\sqrt{2^{n+1}}}\sum_{y}(-1)^{a\cdot y}\left[1+(-1)^{s\cdot y}\right]|y\rangle\,|f(a)\rangle = \frac{1}{\sqrt{2^{n-1}}}\sum_{s\cdot y=0}(-1)^{a\cdot y}\,|y\rangle\,|f(a)\rangle$$

where we have used the fact that $(a\oplus s)\cdot y = (a\cdot y)\oplus(s\cdot y)$ and noticed that $1+(-1)^{s\cdot y}$ can have only two values: either 2 (when $s\cdot y=0$) or 0 (when $s\cdot y=1$). Now we measure the first register. The measurement outcome is selected at random from all possible values of y such that $a\cdot y=0$, each occurring with probability $\frac{1}{2^{n-1}}$.

In fact, we do not have to measure the second register. It was a mathematical shortcut; we did it for purely pedagogical purposes, just to simplify calculations. Instead of 'collapsing' the state to just one term in a superposition we can express Eq.(1) as

$$\frac{1}{2^n}\sum_{y,f(a)}\left[(-1)^{a\cdot y}+(-1)^{(a\oplus s)\cdot y}\right]|y\rangle\,|f(a)\rangle = \frac{1}{2^n}\sum_{y,f(a)}(-1)^{a\cdot y}[1+(-1)^{s\cdot y}]\,|y\rangle\,|f(a)\rangle\,,$$

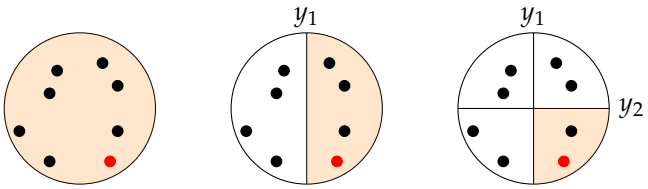
where the summation over $f(a)$ includes all the binary strings in the range of f. The output of the algorithm is then

$$\frac{1}{2^{n-1}}\sum_{s\cdot y=0}|y\rangle\sum_{f(a)}(-1)^{a\cdot y}\,|f(a)\rangle\,,$$

and again, the measurement outcome is selected at random from all possible values of y such that $s \cdot y = 0$. We are not quite done yet. We cannot infer s from a single output y, but one we have found n linearly independent strings $y_1, y_2, \ldots, y_n$ we can solve the n equations

$$s \cdot y_1 = 0, \quad s \cdot y_2 = 0, \quad \ldots \quad s \cdot y_n = 0,$$

to determine a unique value of of s. Thus we run this algorithm repeatedly, each time obtaining another value of y satisfying $s \cdot y = 0$. Each binary string y, assuming that they are independent, allows us to discard half of potential candidates for s,



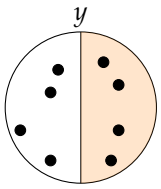
Here, the binary strings which are discarded as candidates for s are shown as black dots on white background. As you can see, we rapidly zoom in on s (the red dot). The probability that $y_1, y_2, \ldots, y_n$ are linearly independent is

$$\left(1 - \frac{1}{2^n}\right)\left(1 - \frac{1}{2^{n-1}}\right) \ldots \left(1 - \frac{1}{2}\right).$$

Indeed, suppose we have k linearly independent binary strings $y_1, \ldots, y_k$. They span a subspace of size 2^k, consisting of all binary strings of the form $b_1 y_1 \oplus b_2 y_2 \oplus \ldots \oplus b_k y_k$, where $b_1, \ldots, b_k \in \{0,1\}$. Now, suppose we get y_{k+1}. It will be independent from $y_1, \ldots, y_k$ only if it lies outside the subspace spanned by $y_1 \ldots, y_k$, which occurs with probability $1 - \frac{2^k}{2^n}$. Now, we can lower bound the expression above as

$$\left[1 - \left(\frac{1}{2^n} + \frac{1}{2^{n-1}} + \ldots + \frac{1}{4}\right)\right] \cdot \frac{1}{2} \geq \frac{1}{4}$$

We conclude that we can determine s with constant probability of error after repeating the algorithm $O(n)$ times. The exponential separation that this algorithm demonstrates between quantum and classical highlights the vast potential of a quantum computer to speed up function evaluation.



$y \cdot x = 1$ on the left, $y \cdot x = 0$ on the right.

Linearly independent here means that no string in the set $\{y_1, y_2, \ldots, y_n\}$ is the bitwise sum of some other strings in this set.

Use the inequality $(1-x)(1-y) = 1 - x - y - xy \geq 1 - (x+y)$ which holds for any $x, y \in (0,1)$

Notes and Exercises

(1) Consider the Boolean function $f : \{0,1\}^n \mapsto \{0,1\}$ defined as $f(x) = a \cdot x$, for some fixed a. Exactly one half of the binary strings $x \in \{0,1\}^n$ give $a \cdot x = 0$ and the other half $a \cdot x = 1$.

DENSITY OPERATORS

ARTUR EKERT

Introduction to Quantum Information Science Lecture 6

We cannot always assign a definite state vector to a quantum system. It may be that the system is part of a composite system that is in an entangled state, or it may be that our knowledge of the preparation of a particular system is insufficient to determine its state. For example, someone may prepare a particle in one of the states $|\psi_1\rangle$, $|\psi_2\rangle$,...,$|\psi_n\rangle$, choosing with probabilities p_1, p_2,...,p_n. Nevertheless, in either case we are able to make statistical predictions about the outcomes of measurements performed on the system using a more general description of quantum states: *density operators*.

We have already mentioned that the existence of entangled states begs an obvious question: if we cannot attribute a state vectors to an individual quantum system then how shall we describe its quantum state? In this lecture we will introduce an alternate description of quantum states that can be applied both to a composite system and to any of its subsystems. Our new mathematical tool is called a density operator. We will start with the density operator as a description of the mixture of quantum states. We will then discuss the partial trace, which is a unique operation that takes care of the reduction of a density operator of a composite system to density operators of its components. If you are an impatient mathematically minded person, who feels more comfortable when things are properly defined right from the beginning, here is your definition:

If we choose a particular basis, operators become matrices. Here I will use both terms – density operators and density matrices – interchangeably.

A density operator ρ on a finite dimensional Hilbert space $\mathcal{H}$ is any non-negative self-adjoint operator with trace equal to one.

A self-adjoint matrix M is called non-negative, or positive semidefinite, if $\langle v|\, M\, |v\rangle \geq 0$ for any vector $|v\rangle$, or if all of its eigenvalues are non-negative, or if here exists a matrix A such that $M = A^\dagger A$ (This is called a Cholesky factorization.)

The rank of a matrix is the number of its non-zero eigenvalues

It follows that ρ can always be diagonalised, the the eigenvalues are all real and nonnegative, and that the eigenvalues sum to one. Moreover, given two density operators ρ_1 and ρ_2, we can always construct another density operator as a convex sum of the two,

$$\rho = p_1\, \rho_1 + p_2\, \rho_2, \quad \text{where } p_1, p_2 \geq 0 \text{ and } p_1 + p_2 = 1.$$

The trace will feature prominently in this lecture. Recall that the trace of a matrix is the sum of its diagonal entries and it is basis independent. Here are some of its properties.

$$\begin{aligned}
&\mathrm{Tr}\,(\alpha A + \beta B) = \alpha \mathrm{Tr}\, A + \beta \mathrm{Tr}\, B \\
&\mathrm{Tr}\,\, |a\rangle\, \langle b| = \langle\, b\,|\, a\,\rangle \\
&\mathrm{Tr}\, ABC = \mathrm{Tr}\, CAB = \mathrm{Tr}\, BCA \\
&\mathrm{Tr}\, A \otimes B = (\mathrm{Tr}\, A)(\mathrm{Tr}\, B)
\end{aligned}$$

A subset of a vector space is said to be convex if the set contains the straight line segment connecting any two points in the set.

You should check that ρ has all the defining properties of a density matrix, i.e. that it is self-adjoint, non-negative and that its trace is one. This means that density operators form a convex set. An important example of a density operator is a rank one projector. Any quantum state that can be described by the state vector $|\psi\rangle$, these are known as pure states, can be also described by the density operator $\rho = |\psi\rangle\, \langle\psi|$. Pure states are the extremal points in the convex set of density operators; they cannot be expressed as a convex sum of other elements in the set. In contrast, all other states, called the mixed states, can be always written as the convex sum of pure states $\sum_i p_i\, |\psi_i\rangle\, \langle\psi_i|$ ($p_i \geq 0$ and $\sum_i p_i = 1$). Now, that we have cleared the mathematical essentials, we will turn to physical applications.

1.1. **Mixtures.** Let us start with probability distributions over state vectors. Suppose Alice prepares a quantum system and hands it over to Bob who subsequently measures observable M. If Alice's preparation is described by a state vector $|\psi\rangle$, then, quantum theory declares, the average value of any observable M is given by $\langle\psi|\, M\, |\psi\rangle$, which can be also written as

$$\langle M\rangle = \mathrm{Tr}\, M\, |\psi\rangle\, \langle\psi|\,.$$

If M is one of the orthogonal projectors P_k describing the measurement then the average $\langle P_k\rangle$ is the probability of the outcome k associated with this projector.

This way of expressing the average value makes a clear separation between the contributions from the state preparation and from the choice of the measurement. We have two operators under the trace, one of them, $|\psi\rangle\, \langle\psi|$, describes the state

preparation and the other one, M, the measurement. Now, suppose Alice prepares the quantum system in one of the states $|\psi_1\rangle, \ldots, |\psi_m\rangle$, choosing state $|\psi_i\rangle$ with probability p_i, and hands the system to Bob without telling him which state was chosen. The possible states $|\psi_i\rangle$ are normalised but need not be orthogonal. We call this situation a mixture of the states $|\psi_i\rangle$, or a mixed state for short. Bob knows the ensemble of states $|\psi_1\rangle, \ldots, |\psi_m\rangle$ and the corresponding probability distribution $p_1, \ldots, p_m$, hence he can calculate $\langle M\rangle$ as

A pure state can be seen as a special case of a mixed state, where all but one the probabilities p_i equal zero.

$$\langle M\rangle = \sum_i p_i \left(\operatorname{Tr} M |\psi_i\rangle \langle\psi_i|\right) = \operatorname{Tr} M \underbrace{\left(\sum_i p_i |\psi_i\rangle \langle\psi_i|\right)}_{\rho} = \operatorname{Tr} M\, \rho.$$

Again, we have two operators under the trace, $\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|$ which pertains to the state preparation and M which describes the measurement. We shall call the operator

Remember, a mixture of states is very different from a superposition of states. A superposition always yields a definite state vector, whereas a mixture does not, and so must be described by a density operator.

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i| \tag{1}$$

the density operator, for it has all the defining properties of the density operator (the convex sum of rank one projectors). It depends on the constituent states $|\psi_i\rangle$ and their probabilities, and it describes our ignorance about the state preparation. Once we have ρ we can make statistical predictions; for any observable M we have,

$$\langle M\rangle = \operatorname{Tr} M\, \rho. \tag{2}$$

We see that the exact composition of the mixture does not enter this formula. For computing the statistics associated with any observable property of a system, all that matters is the density operator itself, and not its decomposition into the mixture of states. This is important because any given density operator, with a remarkable exception of a pure state, can arise from many different mixtures of pure states. Consider, for example, the following three scenarios.

Scenario 1. Alice flips a fair coin. If the result is HEADS she prepares the qubit in the state $|0\rangle$, and if the result is TAILS she prepares the qubit in the state $|1\rangle$. She gives Bob the qubit without revealing the result of the coin-flip. Bob's knowledge of the qubit is described by the density matrix

$$\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

Scenario 2. Suppose Alice flips a fair coin, as before, but now if the result is HEADS she prepares the qubit in the state $|\bar{0}\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle$, and if the result is TAILS she prepares the qubit in the state $|\bar{1}\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle$. Bob's knowledge of the qubit is now described by the density matrix

$$\frac{1}{2}|\bar{0}\rangle\langle\bar{0}| + \frac{1}{2}|\bar{1}\rangle\langle\bar{1}| = \frac{1}{2}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

Scenario 3. Suppose Alice picks up any pair of orthogonal states of a qubit and then flips the coin to chose one of them. Any two orthonormal states of a qubit, $|u_1\rangle$, $|u_2\rangle$, form a complete basis, so the mixture $\frac{1}{2}|u_1\rangle\langle u_1| + \frac{1}{2}|u_2\rangle\langle u_2|$ gives $\frac{1}{2}\mathbb{1}$.

As you can see, these three different preparations yield precisely the same density matrix and hence they are statistically indistinguishable. In general, two different mixtures can be distinguished (in a statistical sense) if and only if they yield different density matrices. In fact, the optimal way of distinguishing quantum states with different density operators is still an active area of research.

1.2. **Few instructive examples and less instructive remarks.**

(1) The density matrix corresponding to the state vector $|\psi\rangle$ is the rank one projector $|\psi\rangle\langle\psi|$. Observe that there is no phase ambiguity for $|\psi\rangle \to e^{i\phi}|\psi\rangle$ leaves the density matrix unchanged, and each $|\psi\rangle$ gives rise to a distinct density matrix.

(2) If Alice prepares a qubit in the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ the corresponding density matrix is the projector

$$|\psi\rangle\langle\psi| = \begin{bmatrix} |\alpha|^2 & \alpha\beta^\star \\ \alpha^\star\beta & |\beta|^2 \end{bmatrix}.$$

(3) You are given a qubit and you are told that it was prepared either in state $|0\rangle$ with probability $|\alpha|^2$ or in state $|1\rangle$ with probability $|\beta|^2$. In this case all you can say is that your qubit is in a mixed state described by the density matrix

$$|\alpha|^2|0\rangle\langle 0| + |\beta|^2|1\rangle\langle 1| = \begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}.$$

Diagonal density matrices correspond to classical probability distributions on the set of basis vectors.

Suppose you want to distinguish between preparations described by the density matrices in the examples (2) and (3). Assume that you are given sufficiently many identically prepared qubits described either by the density matrix in the example (2) or by the density matrix in the example (3). Which of the two measurements would you choose: the measurement in the standard basis $\{|0\rangle, |1\rangle\}$ or the measurement in the basis $\{|\psi\rangle, |\psi_\perp\rangle\}$? One of the two measurements is completely useless. Which one and why?

(4) In general, the diagonal entries of a density matrix describe the probability distributions on the set of basis vectors. They must add up to one, which is why the trace of any density matrix is one. The off-diagonal elements, often called coherences, signal departure from the classical probability distribution and quantify the degree to which a quantum system can interfere (we will discuss this in detail later on). The process in which off-diagonal entries, the parameter ϵ in the matrices below, get zeroed out is called decoherence.

$$\begin{bmatrix} |\alpha|^2 & \alpha\beta^\star \\ \alpha^\star\beta & |\beta|^2 \end{bmatrix} \to \begin{bmatrix} |\alpha|^2 & \epsilon \\ \epsilon^\star & |\beta|^2 \end{bmatrix} \to \begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}$$

For for $\epsilon = \alpha\beta^\star$ we have a pure quantum state (full interference capability) and for $\epsilon = 0$ we have a classical probability distribution over the standard basis (no interference capability).

(5) Suppose it is equally likely that your qubit was prepared either in state $\alpha|0\rangle + \beta|1\rangle$ or in state $\alpha|0\rangle - \beta|1\rangle$. It means that your qubit is in a mixed state described by the density matrix

$$\frac{1}{2}\begin{bmatrix} |\alpha|^2 & \alpha\beta^\star \\ \alpha^\star\beta & |\beta|^2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} |\alpha|^2 & -\alpha\beta^\star \\ -\alpha^\star\beta & |\beta|^2 \end{bmatrix} = \begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}.$$

You cannot tell the difference between the equally weighted mixture of $\alpha|0\rangle \pm \beta|1\rangle$ and a mixture of $|0\rangle$ and $|1\rangle$ with probabilities $|\alpha|^2$ and $|\beta|^2$, respectively.

(6) For any density matrix ρ the most natural mixture that yields ρ is its spectral decomposition, $\rho = \sum_i p_i |u_i\rangle\langle u_i|$, with eigenvectors $|u_i\rangle$ and eigenvalues p_i.

(7) If the states $|u_1\rangle, \ldots, |u_m\rangle$ forms an orthonormal basis, and each occurs with equal probability $1/m$, then the resulting density matrix is proportional to the identity,

$$\frac{1}{m}\sum_{i=1}^{m} |\psi_i\rangle\langle\psi_i| = \frac{1}{m}\mathbb{1}.$$

This is called the maximally mixed state. For qubits, any pair of orthogonal states taken with equal probabilities gives the maximally mixed state $\frac{1}{2}\mathbb{1}$. In the maximally mixed states outcomes of *any* measurement are completely random.

(8) It is often convenient to write density operators in terms of projectors on states which are not normalised, incorporating the probabilities into the length of the state vector,

$$\rho = \sum_i |\tilde{\psi}_i\rangle \langle\tilde{\psi}_i| \,,$$

where $|\tilde{\psi}_i\rangle = \sqrt{p_i}\,|\psi_i\rangle$, that is, $p_i = \langle\, \tilde{\psi}_i \,|\, \tilde{\psi}_i \,\rangle$. This form is more compact but you have to remember that the state vectors are not normalised. I tend to mark such states with the tilde, e.g. $|\tilde{\psi}\rangle$, but you may have your own way to remember.

1.3. **Mixed states of a qubit and the Bloch sphere.** There is an elegant way to visualise the set of density operators for a qubit. The most general Hermitian 2×2 matrix has four real parameters and can be expanded in the basis composed of the identity and the Pauli matrices, $\{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}$. Since the Pauli matrices are traceless, the coefficient of $\mathbb{1}$ in the expansion of a density matrix ρ must be $\frac{1}{2}$, so that $\operatorname{Tr}\rho = 1$. Thus, ρ may be expressed as

Physicists usually refer to the Bloch ball as the Bloch sphere, although it is really a ball, not a sphere.

$$\begin{aligned}\rho &= \frac{1}{2}\left(\mathbb{1} + \vec{s}\cdot\vec{\sigma}\right) = \frac{1}{2}\left(\mathbb{1} + s_x\sigma_x + s_y\sigma_y + s_z\sigma_z\right) \\ &= \frac{1}{2}\begin{bmatrix} 1+s_z & s_x - is_y \\ s_x + is_y & 1-s_z \end{bmatrix}. \end{aligned} \tag{3}$$

The vector $\vec{s}$ is called the Bloch vector for the density operator ρ. Any real Bloch vector $\vec{s}$ defines a trace one Hermitian operator ρ, but in order for ρ to be a density operator it must also be non-negative. Which Bloch vectors yield legitimate density operators? Let us compute the eigenvalues of ρ. The sum of the two eigenvalues of ρ is, of course, equal to one ($\operatorname{Tr}\rho = 1$) and the product is equal to the determinant of ρ, which can be computed from the matrix form above, $\det\rho = \frac{1}{4}(1-s^2) = \frac{1}{2}(1+s)\frac{1}{2}(1+s)$, where $s = |\vec{s}|$. It follows that the two eigenvalues of ρ are $\frac{1}{2}(1 \pm s)$. They have to be non-negative thus s, the length of the Bloch vector cannot exceed one. We can now visualise the convex set of 2×2 density matrices as a unit ball in the three-dimensional Euclidean space. The extremal points, which represent pure states, are the points on the boundary, that is on the surface of the ball ($s = 1$). The maximally mixed state $\mathbb{1}/2$ corresponds to $s = 0$, the centre of the ball. The length of the Bloch vector s can be thought of as a "purity" of a state.

One might hope that there is an equally nice visualisation of the density operators in higher dimensions, unfortunately there isn't.

1.4. **Subsystems of entangled systems.** I have already trumpeted that one of the most important features of the density operator formalism is its ability to describe the quantum state of a subsystem of a composite system. Let me now show you how it works. Given a quantum state of the composite system $\mathcal{AB}$, described by some density operator ρ^{AB}, we obtain reduced density operators ρ^A and ρ^B of subsystems $\mathcal{A}$ and $\mathcal{B}$, respectively by the partial trace.

$$\rho^{AB} \longrightarrow \underbrace{\rho_A = \operatorname{Tr}_B \rho^{AB}}_{\text{partial trace over B}} \qquad \rho^{AB} \longrightarrow \underbrace{\rho_B = \operatorname{Tr}_A \rho^{AB}}_{\text{partial trace over A}}$$

We define the partial trace over $\mathcal{B}$, or $\mathcal{A}$, first on a tensor product of two operators $A \otimes B$ as

$$\operatorname{Tr}_B A \otimes B = A\,(\operatorname{Tr} B), \quad \operatorname{Tr}_A A \otimes B = (\operatorname{Tr} A)\,B,$$

and then extend to any operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ by linearity. Here is a simple example. Suppose a composite system $\mathcal{AB}$ is in a pure entangled state, which we can always write as

$$|\psi_{AB}\rangle = \sum_i c_i\,|a_i\rangle \otimes |b_i\rangle \,, \tag{4}$$

where $|a_i\rangle$ and $|b_j\rangle$ are two orthonormal bases (e.g. the Schmidt bases), and $\sum_i |c_i|^2 = 1$ due to the normalisation. The corresponding density operator of the composite

system is the projector $\rho^{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$, which we can write as

$$\rho^{AB} = |\psi_{AB}\rangle\langle\psi_{AB}| = \sum_{ij} c_i c_j^\star |a_i\rangle\langle a_j| \otimes \underbrace{|b_i\rangle\langle b_j|}_{\text{we will trace this part only}}$$

Let us compute the reduced density operator ρ^A by taking the partial trace over $\mathcal{B}$,

$$\begin{aligned}\rho_A = \mathrm{Tr}_B \rho^{AB} &= \mathrm{Tr}_B |\psi_{AB}\rangle\langle\psi_{AB}| = \mathrm{Tr}_B \sum_{ij} c_i c_j^\star |a_i\rangle\langle a_j| \otimes |b_i\rangle\langle b_j| \\ &= \sum_{ij} c_i c_j^\star |a_i\rangle\langle a_j| \,(\mathrm{Tr}\,|b_i\rangle\langle b_j|) = \sum_{ij} c_i c_j^\star |a_i\rangle\langle a_j| \underbrace{\langle\, b_i \,|\, b_j \,\rangle}_{\delta_{ij}} \\ &= \sum_i |c_i|^2 |a_i\rangle\langle a_i|\,,\end{aligned}$$

where we have used $\mathrm{Tr}\,|b_i\rangle\langle b_j| = \langle\, b_j \,|\, b_i \,\rangle = \delta_{ij}$. In the $|a_i\rangle$ basis the reduced density matrix ρ^A is diagonal, with entries $p_i = |c_i|^2$. We can also take the partial trace over $\mathcal{A}$ and obtain $\rho_B = \sum_i |c_i|^2 |b_i\rangle\langle b_i|$. In particular, for the maximally entangled states in the $d \times d$ dimensional Hilbert spaces $\mathcal{H}_A \otimes \mathcal{H}_B$

$$|\psi_{AB}\rangle = \frac{1}{\sqrt{d}} \sum_i^d |a_i\rangle |b_i\rangle\,, \tag{5}$$

the reduced density operators, ρ_A and ρ_B, are the maximally mixed states, $\frac{1}{d}\mathbb{1}$. It follows that the quantum states of individual qubits in any of the Bell states are maximally mixed, that is, their density matrix is $\frac{1}{2}\mathbb{1}$. The state such as

$$\frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)$$

guarantees perfect correlations when each qubit is measured in the standard basis: the two equally likely outcomes are (0 and 0) or (1 and 1), but, any single qubit outcome, be it 0 or 1 or anything else, is completely random.

1.5. **Partial trace revisited.** If you are given a matrix you calculate the trace by summing its diagonal entries. How about the partial trace? Suppose someone writes down for you a density matrix of two qubits in the standard basis, $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, and asks you to find the reduced density matrices of the individual qubits. The tensor product structure of this 4×4 matrix means that it is has a block form,

$$\rho^{AB} = \left[\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right],$$

where P, Q, R, S are 2×2 sized sub-matrices. The two partial traces can then be evaluated as

$$\rho_A = \mathrm{Tr}_B\, \rho^{AB} = \left[\begin{array}{c|c} \mathrm{Tr}\, P & \mathrm{Tr}\, Q \\ \hline \mathrm{Tr}\, R & \mathrm{Tr}\, S \end{array}\right], \quad \rho_B = \mathrm{Tr}_A\, \rho^{AB} = P + S.$$

Take any of the Bell states, write explicitly its 4×4 density matrix and then trace over each qubit. In each case you should get the maximally mixed state.

The same holds for general ρ^{AB} on any $\mathcal{H}_A \otimes \mathcal{H}_B$ with corresponding block form ($m \times m$ blocks of $n \times n$ sized sub-matrices, where m and n are the dimensions of $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively).

1.6. **Mixtures and subsystems.** We have used the density operators to describe two distinct situations: the statistical properties of the mixtures of states and the statistical properties of subsystems of composite systems. In order to see the relationship between the two, consider a joint state of a bipartite system $\mathcal{AB}$, written in a product basis in $\mathcal{H}_A \otimes \mathcal{H}_B$ as

$$|\psi_{AB}\rangle = \sum_{ij} c_{ij} |a_i\rangle \otimes |b_j\rangle = \sum_{j=1} |\tilde{\psi}_j\rangle |b_j\rangle = \sum_{j=1} \sqrt{p_j}\, |\psi_j\rangle |b_j\rangle\,. \tag{6}$$

where the $|\tilde{\psi}_j\rangle = \sum_i c_{ij} |a_i\rangle = \sqrt{p_j} |\psi_j\rangle$ and vectors $|\psi_j\rangle$ are the normalised versions of $|\tilde{\psi}_j\rangle$. Note that $p_j = \langle \tilde{\psi}_j | \tilde{\psi}_j \rangle$. The partial trace over $\mathcal{B}$ gives the reduced density operator of subsystem $\mathcal{A}$,

$$\begin{aligned}\rho^A =& \mathrm{Tr}_B \sum_{ij} |\tilde{\psi}_i\rangle \langle\tilde{\psi}_j| \otimes |b_i\rangle \langle b_j| = \sum_{ij} |\tilde{\psi}_i\rangle \langle\tilde{\psi}_j| \, (\mathrm{Tr} \, |b_i\rangle \langle b_j|) \\ =& \sum_{ij} |\tilde{\psi}_i\rangle \langle\tilde{\psi}_j| \langle b_j | b_i \rangle = \sum_i |\tilde{\psi}_i\rangle \langle\tilde{\psi}_i| = \sum_i p_i |\psi_i\rangle \langle\psi_i| . \end{aligned} \tag{7}$$

Now, let us see how ρ^A can be understood in terms of mixtures. Let us place subsystems $\mathcal{A}$ and $\mathcal{B}$ in separate labs, run by Alice and Bob, respectively. When Bob measures part $\mathcal{B}$ in the $|b_j\rangle$ basis and obtains result k, which happens with the probability p_k, he prepares subsystem $\mathcal{A}$ in the state $|\psi_k\rangle$,

$$\sum_{i=1} \sqrt{p_j} |\psi_i\rangle |b_i\rangle \stackrel{\text{outcome k}}{\longrightarrow} |\psi_k\rangle |b_k\rangle .$$

Bob does not communicate the outcome of his measurement thus, from Alice's perspective, Bob prepares a mixture of $|\psi_1\rangle,\ldots,|\psi_m\rangle$, with probabilities $p_1,\ldots,p_m$, which means that Alice, who knows the joint state but not the outcomes of Bob's measurement, may associate density matrix $\rho_A = \sum_i p_i |\psi_i\rangle \langle\psi_i|$ with her subsystem $\mathcal{A}$. This is the same ρ^A which we obtained by the partial trace.

But suppose Bob chooses to measure his subsystem in some other basis. Will it have any impact on Alice's statistical predictions? Measurement in the new basis will result in a different mixture but Alice's density operator will not change. Suppose Bob chooses basis $|d_i\rangle$ for his measurement. Any two orthonormal bases are connected by some unitary transformation, hence we can write $|b_i\rangle = U |d_i\rangle$ for some unitary U. In terms of components $|b_i\rangle = \sum_j U_{ij} |d_j\rangle$. The joint state can now be expressed as

$$\begin{aligned}|\psi_{AB}\rangle = \sum_i |\tilde{\psi}_i\rangle |b_i\rangle &= \sum_i |\tilde{\psi}_i\rangle \left(\sum_j U_{ij} |d_j\rangle \right) \\ &= \sum_j \underbrace{\left(\sum_i U_{ij} |\tilde{\psi}_i\rangle \right)}_{|\tilde{\phi}_j\rangle} |d_j\rangle = \sum_j |\tilde{\phi}_j\rangle |d_j\rangle . \end{aligned} \tag{8}$$

If Bob measures in the $|d_i\rangle$ basis then he generates a new mixture of states $|\phi_1\rangle, \ldots |\phi_m\rangle$, which are the normalised versions of $|\tilde{\phi}_1\rangle, \ldots |\tilde{\phi}_m\rangle$, each $|\phi_k\rangle$ occurring with the probabilities $p_k = \langle \tilde{\phi}_k | \tilde{\phi}_k \rangle$. But this new mixture has exactly the same density operator as the previous one,

U_{ij} are the components of a unitary matrix hence $\sum_k U_{ik} U^\star_{jk} = \delta_{ij}$.

$$\sum_j |\tilde{\phi}_j\rangle \langle\tilde{\phi}_j| = \sum_{ijl} U_{ij} |\tilde{\psi}_i\rangle \langle\tilde{\psi}_l| U^\star_{lj} = \sum_{il} \underbrace{\left(\sum_j U_{ij} U^\star_{lj} \right)}_{\delta_{il}} |\tilde{\psi}_i\rangle \langle\tilde{\psi}_l| = \sum_i |\tilde{\psi}_j\rangle \langle\tilde{\psi}_j| ,$$

which is exactly ρ_A. Does it really matter whether Bob performs the measurement or not? It does not. After all Alice and Bob may be miles away from each other and if any of Bob's actions were to result in something that is physically detectable at the Alice's location that would amount to instantaneous communication between the two of them.

From the operational point of view it does not really matter whether the density operator represents our ignorance of the actual state (mixtures) or provides the only description we can have after discarding one part of an entangled state (partial trace). In the former case the system is in some definite pure state but we do not know which. In contrast, when the density operator arises from tracing out irrelevant, or unavailable, degrees of freedom the individual system cannot be thought to be in some definite state of which we are ignorant. Philosophy aside, the fact

The two interpretations of density operators filled volumes of academic papers. The terms "proper mixtures" and "improper mixtures" are used, mostly by philosophers, to describe the statistical mixture and the partial trace approach, respectively.

that the two interpretations give exactly the same predictions is useful. Switching back and forth between the two pictures often offers additional insights and may even simplify lengthy calculations.

1.7. **Partial trace again.** The partial trace is the only map $\rho^{AB} \to \rho^A$ such that

$$\operatorname{Tr} X\rho^A = \operatorname{Tr}(X \otimes \mathbb{1})\rho^{AB}$$

holds for any observable X acting on $\mathcal{A}$. This condition is about the consistency of statistical predictions. Any observable X on $\mathcal{A}$ can be viewed as an observable $X \otimes \mathbb{1}$ on the composite system $\mathcal{AB}$, where $\mathbb{1}$ is the identity operator acting on $\mathcal{B}$. When constructing ρ^A we had better make sure that for any observable X the average value of X in the state ρ_A is the same as the average value of $X \otimes \mathbb{1}$ in the state ρ^{AB}. This is indeed the case for the partial trace. For example, let us go back to the state in Eq.(6) and assume that Alice measures some observable X on her part of the system. Technically, such an observable can be expressed as $X \otimes \mathbb{1}$, where $\mathbb{1}$ is the identity operator acting on the subsystem $\mathcal{B}$. The expectation value of this observable in the state $|\psi_{AB}\rangle$ is $\operatorname{Tr}(X \otimes \mathbb{1}) |\psi_{AB}\rangle \langle\psi_{AB}|$, that is,

$$\begin{aligned}
\operatorname{Tr}(X \otimes \mathbb{1})\rho^{AB} &= \operatorname{Tr}(X \otimes \mathbb{1}) \left(\sum_{ij} |\tilde{\psi}_i\rangle \langle\tilde{\psi}_j| \otimes |b_i\rangle \langle b_j| \right) \\
&= \sum_{ij} \left[\operatorname{Tr}\left(X |\tilde{\psi}_i\rangle \langle\tilde{\psi}_j|\right)\right] \underbrace{\left[\operatorname{Tr}\left(|b_i\rangle \langle b_j|\right)\right]}_{\delta_{ij}} \\
&= \sum_i \operatorname{Tr} X |\tilde{\psi}_i\rangle \langle\tilde{\psi}_i| = \operatorname{Tr} X \underbrace{\sum_i p_i |\psi_i\rangle \langle\psi_i|}_{\rho^A = \operatorname{Tr}_B \rho^{AB}} \\
&= \operatorname{Tr} X\rho^A, \qquad (9)
\end{aligned}$$

as required.

One can repeat the same argument for $\rho^{AB} \to \rho^B$.

Similarly, the partial trace is a unique map $\rho^{AB} \to \rho^B$ such that ρ^B satisfies $\operatorname{Tr} Y\rho^B = \operatorname{Tr}(1 \otimes Y)\rho^{AB}$ for any observable Y on $\mathcal{B}$.

TO BE COMPLETED: the uniqueness of the partial trace, for now see Nielsen& Chuang Box 2.6.

Notes and Exercises

(1) Show that an arbitrary mixed state ρ can be represented as the partial trace $\operatorname{Tr} |\psi\rangle \langle\psi|$ of a pure state of a larger system. Such $|\psi\rangle$ is called a purification of ρ.
(2) Show that purification is unique up to unitary equivalence. Let $|\psi_1\rangle$ and $|\psi_2\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_B$ be two pure states such that $\operatorname{Tr}_B |\psi_1\rangle \langle\psi_1| = \operatorname{Tr}_B |\psi_2\rangle \langle\psi_2|$. Show that $|\psi_1\rangle = \mathbb{1} \otimes U |\psi_2\rangle$ for some unitary operator U on $\mathcal{H}_B$.
(3) Two qubits are in the state described by the density operator $\rho = \rho_A \otimes \rho_B$. What is the partial trace of ρ over each qubit?
(4) Write the density matrix of two qubits corresponding to the mixture of the Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ with probability $\frac{1}{2}$ and the maximally mixed state of two qubits (4×4 matrix in $\mathcal{H}_A \otimes \mathcal{H}_B$) with probability $\frac{1}{2}$.
(5) Trace norm (to be completed)
(6) How to distinguish between two different density operators (to be completed)

COMPLETELY POSITIVE TRACE PRESERVING MAPS

ARTUR EKERT

Introduction to Quantum Information Science
Lecture 7

Quantum evolution of isolated systems is unitary but their subsystems may evolve in a more complicated way. In this lecture we will discuss linear transformations, often called quantum channels, which map density operators into density operators and elaborate on a subtle difference between positive maps and completely positive maps (physically admissible), so that you know what is physically possible and what is not.

We have spent some time discussing unitary evolutions: state $|\psi\rangle$ evolves into another state $U|\psi\rangle$ according to some unitary operator U. In terms of density matrices, a unitary operation U applied to a pure state $|\psi\rangle\langle\psi|$ results in the density matrix $U|\psi\rangle\langle\psi|U^\dagger$ (to be consistent with $(U|\psi\rangle)^\dagger = \langle\psi|U^\dagger$). Now, any density operator ρ can be written as a mixture of pure states $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, thus applying U to the mixed state ρ results in the state

$$\rho \longrightarrow U\rho U^\dagger, \qquad U^\dagger U = UU^\dagger = \mathbb{1}.$$

We can go further and consider a probability distribution on a finite set of unitary operations $U_1, \ldots, U_m$. Somebody randomly chooses U_k according to the probability distribution $p_1 \ldots p_k \ldots p_m$ and applies it to ρ without telling you the value of k. This is very reminiscent of our discussion of the mixtures of pure states. The resulting transformation can be written as

$$\rho \longrightarrow \sum_i p_i\, U_i \rho U_i^\dagger = \sum_i \underbrace{\sqrt{p_i}U_i}_{A_i}\ \rho\ \underbrace{U_i^\dagger\sqrt{p_i}}_{A_i^\dagger} = \sum_i A_i \rho A_i^\dagger, \qquad \sum_i A_i^\dagger A_i = \sum_i A_i A_i^\dagger = \mathbb{1}.$$

Let us now make a leap and consider a much more general set of possible operations. We conjecture that any linear map T, which can be written as

$$\rho \longrightarrow T(\rho) = \sum_i A_i \rho A_i^\dagger, \qquad \sum_i A_i^\dagger A_i = \mathbb{1},$$

This way of writing quantum evolutions is known as the Kraus representation or the operator sum representation. The operators A_i are sometimes called Kraus operators. Karl Kraus (1938–1988) was a German physicist.

for some collection of matrices $A_1, \ldots, A_m$, satisfying $\sum_i A^\dagger A = \mathbb{1}$, represents an operation that can, at least in principle, be physically implemented. Such physically admissible quantum operations appear in the technical literature under various names, e.g. superoperators, quantum channels or completely positive trace preserving maps. We can easily check that $T(\rho)$ is Hermitian, has trace one,

$$\mathrm{Tr}\, T(\rho) = \mathrm{Tr} \sum_i A_i \rho A_i^\dagger = \mathrm{Tr} \underbrace{\left(\sum_i A_i^\dagger A_i\right)}_{\mathbb{1}} \rho = \mathrm{Tr}\,\rho = 1,$$

and is non-negative

$$\langle v|\, T(\rho)\, |v\rangle = \sum_i \langle v|\, A_i \rho A_i^\dagger\, |v\rangle = \sum_i \langle v_i|\, \rho\, |v_i\rangle \geq 0, \qquad |v_i\rangle = A_i^\dagger |v\rangle$$

Example Consider a single qubit and the operation T specified by the following two Kraus operators

$$A_1 = |0\rangle\langle 0|, \quad A_2 = |1\rangle\langle 1|.$$

First let us check that this is a valid admissible operation,

$$A_1^\dagger A_1 + A_2^\dagger A_2 = |0\rangle\langle 0|0\rangle\langle 0| + |1\rangle\langle 1|1\rangle\langle 1| = |0\rangle\langle 0| + |1\rangle\langle 1| = \mathbb{1}.$$

It satisfies the required property, hence it is admissible. Let us see what it does to the state $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$.

$$\begin{aligned} T(|\psi\rangle\,\langle\psi|) &= A_1\,|\psi\rangle\,\langle\psi|\,A_1^\dagger + A_2\,|\psi\rangle\,\langle\psi|\,A_2^\dagger \\ &= |0\rangle\langle 0\,|\,\psi\,\rangle\langle\,\psi\,|\,0\,\rangle\langle 0| + |1\rangle\langle\,1\,|\,\psi\,\rangle\langle\,\psi\,|\,1\,\rangle\langle 1| \\ &= |\alpha|^2\,|0\rangle\,\langle 0| + |\beta|^2\,|1\rangle\,\langle 1|\,. \end{aligned} \tag{1}$$

In terms of matrices

$$\begin{bmatrix} |\alpha|^2 & \alpha\beta^\star \\ \alpha^\star\beta & |\beta|^2 \end{bmatrix} \xrightarrow{\;T\;} \begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix},$$

The action of this map is equivalent to measuring the qubit in the standard basis and forgetting the result.

which you should recognise as the decoherence in the standard basis (the off-diagonal elements, called coherences, disappear).

Please note that preserving the trace requires only $\sum_i A_i^\dagger A_i = \mathbb{1}$, that is, we do not require $\sum_i A_i A_i^\dagger = \mathbb{1}$. We do not even require that A_i are square matrices. In general, admissible operations do not need to preserve the sizes of quantum systems. Recall, for example, the partial trace, which maps density matrices on tensor product spaces into smaller density matrices on the constituent spaces. The partial trace is certainly a physically admissible operation for it amounts to discarding one of the subsystems.

1.1. **Mathematical pitstop.** Let us add some notation that may help us to navigate through all these operations on operators. Given a pair of Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ we will denote the set of (bounded) linear operators from $\mathcal{H}_1$ to $\mathcal{H}_2$ as $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$. The shorthand $\mathcal{B}(\mathcal{H})$ is used for $\mathcal{B}(\mathcal{H}, \mathcal{H})$. For example $\rho \in \mathcal{B}(\mathcal{H})$ is the density operator on $\mathcal{H}$. Now, if we have a collection of operators (matrices) $A_1, \ldots, A_m \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ that satisfy $\sum_i A_i^\dagger A_i = \mathbb{1}$ (the identity is in $\mathcal{B}(\mathcal{H}_1)$, then the admissible operation T, defined by

Every operator between finite dimensional vector spaces (which includes all finite matrices) is bounded.

$$T(\rho) = \sum_i^m A_1 \rho A_i^\dagger$$

maps elements of $\mathcal{B}(\mathcal{H}_1)$ to elements of $\mathcal{B}(\mathcal{H}_2)$,

$$T : \mathcal{B}(\mathcal{H}_1) \longrightarrow \mathcal{B}(\mathcal{H}_2).$$

The composition of admissible operations is an admissible operation. Indeed, let $T = \sum_i A_i \cdot A_i^\dagger$ and $S = \sum_j B_j \cdot B_j^\dagger$ then $ST = \sum_{ij} (B_j A_i) \cdot (B_j A_i)^\dagger$ and

$$\sum_{ij} (B_j A_i)^\dagger (B_j A_i) = \sum_i A_i^\dagger \left(\sum_j B_j^\dagger B_j \right) A_i = \sum_i A_i^\dagger A_i = \mathbb{1}.$$

We shall now discuss two important operations, namely, adding and discarding subsystems. You need to pay attention to the dimensions of the Hilbert spaces involved, for they are different. Given a composite system $\mathcal{AB}$ we can discard subsystem $\mathcal{B}$, which is described by a map (the partial trace)

$$\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \longrightarrow \mathcal{B}(\mathcal{H}_A).$$

Conversely, given system $\mathcal{A}$ we can bring in another system $\mathcal{B}$ to form a combined system $\mathcal{AB}$. This will be described by a map (the isometric embedding)

$$\mathcal{B}(\mathcal{H}_A) \longrightarrow \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B).$$

These are, clearly, physically admissible operations which we will write in the operator sum decomposition.

What happens when you trash a qubit

1.2. **Discarding stuff.** Suppose we have two qubits, $\mathcal{A}$ and $\mathcal{B}$, in some state described by the density operator $\rho^{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$. We will consider the admissible operation that corresponds to discarding the second qubit (qubit $\mathcal{B}$),

$$\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \longrightarrow \mathcal{B}(\mathcal{H}_A).$$

We already know this operation, it is the partial trace over $\mathcal{B}$, but now we want to write it in the operator sum decomposition. The corresponding Kraus operators A_i must come from the set $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_A)$, i.e. they must be 2×4 matrices. Indeed, they are:

$$A_1 = \mathbb{1} \otimes \langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and

$$A_2 = \mathbb{1} \otimes \langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We check that they satisfy $A_1^\dagger A_1 + A_2^\dagger A_2 = \mathbb{1}$,

$$(\mathbb{1} \otimes |0\rangle)(\mathbb{1} \otimes \langle 0|) + (\mathbb{1} \otimes |1\rangle)(\mathbb{1} \otimes \langle 1|) = \mathbb{1} \otimes (|0\rangle\langle 0| + |1\rangle\langle 1|) = \mathbb{1} \otimes \mathbb{1}$$

The identity operator acts on $\mathcal{H}_A \otimes \mathcal{H}_B$. The map takes any tensor product operators of the form $X \otimes Y \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ to

$$A_1(X \otimes Y)A_1^\dagger + A_2(X \otimes Y)A_2^\dagger = X\langle 0|Y|0\rangle + X\langle 1|Y|1\rangle = X\mathrm{Tr}\, Y,$$

which is the definition of the partial trace over $\mathcal{B}$. In general, given a bipartite quantum system $\mathcal{AB}$ with the orthonormal basis $|b_i\rangle$ in $\mathcal{B}$, the Kraus operators corresponding to the partial trace over $\mathcal{B}$ are $A_i = \mathbb{1} \otimes \langle b_i|$. You could replace the vectors $|b_i\rangle$ (and their corresponding bra vectors) by any other orthonormal basis in $\mathcal{H}_B$ without changing the operation. Another way of expressing the partial trace, which makes this fact apparent, is $\mathrm{Tr}_B = \mathbb{1} \otimes \mathrm{Tr}$. Needless to say, we can repeat the same arguments for Tr_A, the partial trace over $\mathcal{A}$.

What happens when you find a qubit.

1.3. **Adding stuff.** We can discard a quantum system, but we can also bring one in. Suppose you have a qubit in some state ρ^A and you bring another one in state ρ^B, so that the joint state of the two qubits is $\rho^{AB} = \rho^A \otimes \rho^B$. This is clearly an admissible operation of the form

$$\mathcal{B}(\mathcal{H}_A) \longrightarrow \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B).$$

In order to find the operator sum decomposition we start with $V \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_A \otimes \mathcal{H}_B)$ that maps any state $|\psi\rangle \in \mathcal{H}_A$ to state $|\psi\rangle \otimes |b\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, for some fixed $|b\rangle$. This operation is called the isometric embedding and, as you can easily check, it satisfies $V^\dagger V = \mathbb{1}$. Thus

Isometric, because it preserves the inner product.

$$V|\psi\rangle\langle\psi|V^\dagger = |\psi\rangle\langle\psi| \otimes |b\rangle\langle b|,$$

and in general, for any ρ^A (any ρ^A can always be written as a mixture of pure states),

$$V\rho^A V^\dagger = \rho^A \otimes |b\rangle\langle b|, \qquad V^\dagger V = \mathbb{1}.$$

Choosing different isometric embeddings V_i with prescribed probabilities p_i, we get

$$\sum_i p_i V_i \rho^A V_i^\dagger = \rho^A \otimes \rho^B.$$

for any ρ^B. This expression can be also written as $\sum_i A_i \rho^A A_i^\dagger$, where $A_i = \sqrt{p_i} V_i$, and $\sum A_i^\dagger A_i = \mathbb{1}$.

1.4. **Quantum operations.** We now postulate that any physically admissible operation is a composition of an arbitrary number of isometric embeddings (adding stuff), unitary transformations and partial tracings (discarding stuff). A typical sequence may look like this:

(1) **Embedding**. Given a physical system described by ρ, we append an auxiliary quantum system in some pure state, usually written as $|0\rangle$,

$$\rho \mapsto \rho \otimes |0\rangle \langle 0| .$$

(2) **Unitary evolution**. The two systems interact and evolve according to some unitary operator U,

$$\rho \otimes |0\rangle \langle 0| \mapsto U \ (\rho \otimes |0\rangle \langle 0|) \, U^\dagger .$$

(3) **Partial trace**. We discard the auxiliary system

$$U \ (\rho \otimes |0\rangle \langle 0|) \, U^\dagger \mapsto \mathrm{Tr}_{\,\mathrm{aux}} \left(U \ (\rho \otimes |0\rangle \langle 0|) \, U^\dagger \right)$$

All together,

$$\rho \mapsto \rho' = \mathrm{Tr}_{\,\mathrm{aux}} [(U(\rho \otimes |0\rangle \langle 0|) U^\dagger]. \tag{2}$$

Since, as we have demonstrated, each of these operations can be represented in terms of the Kraus operators, we conclude that a quantum operation is physically admissible if admits an operator sum representation. The reverse is also true. Given a map which can be represented as the sum of Kraus operators we can always write it as in Eq.(2). (to be completed).

1.5. **Complete positivity.** One may take a more mathematical approach and argue that any linear superoperator T that maps density operators to density operators should represent an admissible physical operation. This leads to two requirements:

- T must be a trace preserving map, so that $\mathrm{Tr}\, T(\rho) = \mathrm{Tr}\, \rho$ for all operators ρ.
- T must be a positive map, in the sense that it maps non-negative operators to non-negative operators.

It turns out, however, that quantum operations are more subtle. The two conditions are clearly necessary, but they are not sufficient. The reason is that no quantum system is alone in the Universe. Consider another, an independent, system, whose state evolves according to the identity map $\mathbb{1}$. If the first system evolved according to T, then the composite system would evolve according to $T \otimes \mathbb{1}$, defined by its action on product operators

$$T \otimes \mathbb{1}(A \otimes B) = T(A) \otimes B.$$

Even though T and $\mathbb{1}$ are positive operators, the combination $T \otimes \mathbb{1}$ may not be! There is an entangled state of the two systems which is mapped by $T \otimes \mathbb{1}$ to an operator that is not non-negative, and hence not a legitimate density operator. In order to see this take, for example, the transpose operation on a single qubit, $T(|i\rangle\langle j|) = |j\rangle\langle i|$, $(i, j = 0, 1)$. In the matrix form,

$$\rho = \begin{bmatrix} a & c \\ c^\star & b \end{bmatrix} \xrightarrow{T} \rho' = \begin{bmatrix} a & c^\star \\ c & b \end{bmatrix}.$$

It certainly preserves both trace and positivity, and the result $\rho' = \rho^T$ is a density matrix. However, if this qubit is part of a two qubit system initially in the entangled state $\frac{1}{\sqrt{2}}(|0\rangle\,|0\rangle + |1\rangle\,|1\rangle)$, and the transpose is applied only to one of the two qubits, say, the first one, then, under the action of the partial transposition $T \otimes \mathbb{1}$, the density matrix of the two qubits evolves as

$$\frac{1}{2} \sum_{ij} |i\rangle\langle j| \otimes |i\rangle\langle j| \longrightarrow \frac{1}{2} \sum_{ij} T(|i\rangle\langle j|) \otimes |i\rangle\langle j| = \frac{1}{2} \sum_{ij} |j\rangle\langle i| \otimes |i\rangle\langle j| ,$$

which can be also written explicitly in the matrix form

$$\frac{1}{2}\begin{bmatrix}1&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&1\end{bmatrix} \xrightarrow{T\otimes\mathbb{1}} \frac{1}{2}\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}.$$

The resulting matrix has a negative eigenvalue so it is not a density matrix and hence it does not represent a quantum state. The transpose is an example of a positive map which is not completely positive. A completely positive map is a map which preserves positivity of the density operator not only of the principal system but also of any extension of this system.

1.6. **Take half of an entangled state.** How can we tell if a given map T is completely positive or not? Pretty much in the same way we discovered that the transpose is not a completely positive map. We simply apply $T\otimes\mathbb{1}$, or $\mathbb{1}\otimes T$, to the maximally entangled state of two subsystems and check if the result is a legal density matrix. There is a very nice isomorphism, known as the Choi-Jamiolkowski isomorphism, between completely positive maps $T:\mathcal{B}(\mathcal{H}_1)\to\mathcal{B}(\mathcal{H}_2)$ and density matrices in $\mathcal{B}(\mathcal{H}_1\otimes\mathcal{H}_2)$, which we shall now describe.

Any linear map T acting on density matrices of some system $\mathcal{B}$ of dimension d can be completely characterised by its action on the d^2 basis matrices $|i\rangle\langle j|$, $T(\rho)=\sum_{ij}\rho_{ij}T(|i\rangle\langle j|)$, where $i,j=1,\ldots,d$. We can then arrange matrices $T(|i\rangle\langle j|)$ into a block matrix $\tilde{T}$, so that the (i,j) entry is the matrix $T(|i\rangle\langle j|)$. The block matrix $\tilde{T}$ is usually called the Choi matrix of T. For example, any linear map T acting on a qubit can be completely characterised by its action on the four basis matrices $|i\rangle\langle i|\,j$, where $i,j=0,1$, and represented as the 4×4 Choi matrix,

This time it will be easier to work with system $\mathcal{B}$ and prepend to it system $\mathcal{A}$. Needless to say, this is just a matter of convenience. Our equations will be less cluttered.

A system of dimension d is, of course, a shortcut for a system with which we associate a Hilbert space of dimension d.

$$\tilde{T}=\frac{1}{2}\left[\begin{array}{c|c}T(|0\rangle\langle0|)&T(|0\rangle\langle1|)\\\hline T(|1\rangle\langle0|))&T(|1\rangle\langle1|)\end{array}\right]. \tag{3}$$

Now, the Choi matrix is also the result of applying $\mathbb{1}\otimes T$ to a maximally entangled state. Given a system $\mathcal{B}$ of dimension d we prepend to it system $\mathcal{A}$ with the same dimension and consider the maximally entangled state of form

$$|\omega\rangle=\frac{1}{\sqrt{d}}\sum_i^d|i\rangle\,|i\rangle\,,\qquad|\omega\rangle\langle\omega|=\frac{1}{d}\sum_{ij}|i\rangle\langle j|\otimes|i\rangle\langle j|\,,$$

where we have used the same symbol $|i\rangle$, $i=1,\ldots,d$ to denote two orthonormal bases for $\mathcal{A}$ and $\mathcal{B}$ (for qubits $i=0,1$). Then we apply $\mathbb{1}\otimes T$,

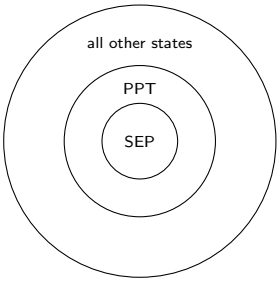
$$(\mathbb{1}\otimes T)\,|\omega\rangle\langle\omega|=\frac{1}{d}\sum_{ij}|i\rangle\langle j|\otimes T(|i\rangle\langle j|)=\tilde{T} \tag{4}$$

and get the corresponding $d^2\times d^2$ Choi matrix. Depending on the map T, the Choi matrix may or may not be a density matrix. If T is a completely positive trace preserving map then, obviously, its extension $\mathbb{1}\otimes T$ is a positive map and $\tilde{T}$ is a legal density operator. Conversely, if the Choi matrix $\tilde{T}$ is a density operator then T is a completely positive map (proof based on representing $\tilde{T}$ as a mixture of pure states $|\psi_i\rangle\langle\psi_i|$ of $\mathcal{AB}$ to be completed).

1.7. **What are the positive maps good for?** Positive maps which are not completely positive are not completely useless. True, they cannot describe any quantum dynamics but still they have useful applications. For example, they can help us to determine if a given state is entangled or not. A quantum state of a bipartite system $\mathcal{AB}$ described by the density matrix ϱ^{AB} is called separable if ϱ^{AB} is of the form

$$\varrho^{AB}=\sum_k p_k\,\rho_k^A\otimes\rho_k^B,$$

where $p_k \geq 0$ and $\sum_{k=1} p_k = 1$. Otherwise ϱ^{AB} is called entangled. If we apply the partial transpose, $\mathbb{1} \otimes T$, to this state it remains separable for, as we have seen, the transposed ρ^B is a legal density matrix.Positive maps, such as the transpose, can be quite deceptive; you have to include other systems in order to detect their unphysical character. In separable states one subsystem does not really know about the existence of the other, hence applying a positive map to one part produces a proper density operator and thus does not reveal the unphysical character of the map. Thus for any separable state ρ we have $(\mathbb{1} \otimes T)\rho \geq 0$. As an example, consider a quantum state of two qubits which is a mixture of the maximally entangled state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and the maximally mixed state, described by the density matrix



$$\rho = p\,|\psi\rangle\langle\psi| + (1-p)\frac{1}{4}\mathbb{1} \otimes \mathbb{1}, \quad p \in [0,1].$$

Apply partial transpose $\mathbb{1} \otimes T$ to this state and check for which values of p the resulting matrix is a density matrix. You should get $p \leq \frac{1}{3}$, which implies that for all p from $\frac{1}{3}$ to 1 the density operator ρ describes an entangled state. Note that the implication "if separable than the partial transpose is positive" does not imply the reverse. There exist entangled states for which the partial transpose is positive, they are known as the entangled PPT states (PPT for positive partial transpose), however, for two qubits the PPT states are exactly the separable states.

Notes and Exercises

(1) Let T on a qubit be defined as,

$$T(\mathbb{1}) = \mathbb{1}, \quad T(\sigma_x) = x\sigma_x, \quad T(\sigma_y) = y\sigma_y, \quad T(\sigma_z) = z\sigma_z,$$

where x, y, z are some real numbers. Using the Choi matrix of T determine the range of x, y, z for which the map T is positive and the range for which it is completely positive.

DECOHERENCE AND ELEMENTS OF QUANTUM ERROR CORRECTION

ARTUR EKERT

In principle we know how to build a quantum computer; we can start with simple quantum logic gates and try to integrate them together into quantum networks. However, if we keep on putting quantum gates together into networks we will quickly run into some serious practical problems. The more interacting qubits are involved, the harder it is to prevent them from getting entangled with the environment. This unwelcome entanglement, also known as decoherence, destroys the interference and the power of quantum computing.

1.1. **Decoherence simplified.** Consider the following qubit-environment interaction,

$$|0\rangle\,|e\rangle \mapsto |0\rangle\,|e_{00}\rangle\,, \quad |1\rangle\,|e\rangle \mapsto |1\rangle\,|e_{11}\rangle$$

where $|e\rangle$, $|e_{00}\rangle$ and $|e_{11}\rangle$ are the states of the environment, which not need to be orthogonal. Let $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ be the initial state of the qubit. The environment is essentially trying to measure the qubit and, as the result, the two get entangled,

$$(\alpha\,|0\rangle + \beta\,|1\rangle)\,|e\rangle \mapsto \alpha\,|0\rangle\,|e_{00}\rangle + \beta\,|1\rangle\,|e_{11}\rangle\,.$$

The reason we use two indices in $|e_{00}\rangle$ and $|e_{11}\rangle$ will become clear in a moment, when we consider more general interaction with the environment

This state can also be written as

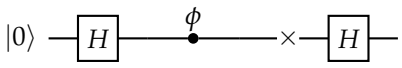
$$\begin{aligned}(\alpha\,|0\rangle + \beta\,|1\rangle)\,|e\rangle \mapsto &(\alpha\,|0\rangle + \beta\,|1\rangle)\,\frac{|e_{00}\rangle + |e_{11}\rangle}{2}\\ &+(\alpha\,|0\rangle - \beta\,|1\rangle)\,\frac{|e_{00}\rangle - |e_{11}\rangle}{2}.\end{aligned}$$

or as

$$|\psi\rangle\,|e\rangle \mapsto \mathbb{1}\,|\psi\rangle\,|e_1\rangle + Z\,|\psi\rangle\,|e_z\rangle\,,$$

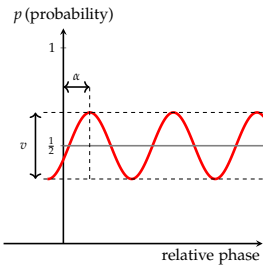
where $|e_1\rangle = \frac{1}{2}(|e_{00}\rangle + |e_{11}\rangle)$ and $|e_z\rangle = \frac{1}{2}(|e_{00}\rangle - |e_{11}\rangle)$. We may interpret this expression by saying that two things can happen to the qubit: nothing $\mathbb{1}$ (first term) or phase-flip Z (second term). This, however, should not be taken literally unless the states of the environment, $|e_1\rangle$ and $|e_z\rangle$ are orthogonal (why?).

1.2. **Decoherence and interference.** Suppose the qubit undergoes the usual interference experiment but in between the two Hadamard gates it is affected by the decoherence (marked by $\times$)



Let us step through the circuit, keeping track of the state of the environment,

$$\begin{aligned}|0\rangle\,|e\rangle &\overset{H}{\mapsto} (|0\rangle + |1\rangle)\,|e\rangle\\ &\overset{\phi}{\mapsto} (|0\rangle + e^{i\phi}\,|1\rangle)\,|e\rangle\\ &\overset{\times}{\mapsto} |0\rangle\,|e_0\rangle + e^{i\phi}\,|1\rangle\,|e_1\rangle\\ &\overset{H}{\mapsto} |0\rangle\,(|e_{00}\rangle + e^{i\phi}\,|e_{11}\rangle) + |1\rangle\,(|e_{00}\rangle - e^{i\phi}\,|e_{11}\rangle).\end{aligned}$$



If we write $\langle\, e_0\,|\,e_1\,\rangle = ve^{i\alpha}$ then the final probabilities of 0 and 1 oscillate with ϕ as

$$\begin{aligned}P_0(\phi) &= \tfrac{1}{2}\,(1 + v\cos(\phi + \alpha))\,,\\ P_1(\phi) &= \tfrac{1}{2}\,(1 - v\cos(\phi + \alpha))\,.\end{aligned}$$

As we can see in the plot, the interference pattern is suppressed by factor v, called visibility. As $v = |\langle e_0 | e_1 \rangle|$ decreases, we lose all the advantages of quantum interference. For example, in Deutsch's algorithm we obtain the correct answer with probability at most $(1+v)/2$. For $\langle e_0 | e_1 \rangle = 0$, the perfect decoherence case, the network outputs 0 or 1 with equal probabilities, *i.e.* it is useless as a computing device.

In terms of density operators, the qubit alone evolves from the pure state $|\psi\rangle\langle\psi|$ to a mixed state, which can be obtained by tracing over the environment,

$$\begin{bmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{bmatrix} \mapsto \begin{bmatrix} |\alpha|^2 & \alpha\beta^*\langle e_{11}|e_{00}\rangle \\ \alpha^*\beta\langle e_{00}|e_{11}\rangle & |\beta|^2 \end{bmatrix}.$$

The off-diagonal elements, originally called "coherences", vanish as $\langle e_0 | e_1 \rangle$ approaches zero. This is why this particular interaction is called decoherence.

It is clear that we want to avoid decoherence, or at least diminish its impact on our computing device. For this we need quantum error correction; we encode the state of a single (logical) qubit across several (physical) qubits.

1.3. **Quantum errors.** The most general qubit-environment interaction,

$$|0\rangle|e\rangle \mapsto |0\rangle|e_0\rangle + |1\rangle|d_0\rangle\,,\ |1\rangle|e\rangle \mapsto |1\rangle|e_1\rangle + |0\rangle|d_1\rangle\,,$$

where the states of the environment are neither normalised nor orthogonal, leads to decoherence

$$\begin{aligned}(\alpha|0\rangle + \beta|1\rangle)|e\rangle \mapsto &(\alpha|0\rangle + \beta|1\rangle)\,\frac{|e_0\rangle + |e_1\rangle}{2} \\ &+(\alpha|0\rangle - \beta|1\rangle)\,\frac{|e_0\rangle - |e_1\rangle}{2} \\ &+(\alpha|1\rangle + \beta|0\rangle)\,\frac{|d_0\rangle + |d_1\rangle}{2} \\ &+(\alpha|1\rangle - \beta|0\rangle)\,\frac{|d_0\rangle - |d_1\rangle}{2}.\end{aligned}$$

Four things can happen to the qubit: nothing, phase-flip, bit-flip or both bit-flip and phase-flip. Essentially we have to deal with two types of quantum errors: bit-flips and phase-flips.

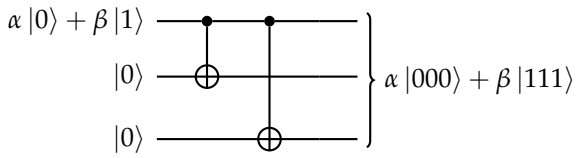
1.4. **Repetition codes.** In order to give a sense of how quantum error correction actually works let us begin with a classical example of a repetition code. Suppose a transmission channel flips each bit in transit with probability p. If this error rate is considered too high then it can be decreased by encoding each bit into, say, three bits,

$$0 \mapsto 000, \quad 1 \mapsto 111.$$

That is, each time we want to send logical 0 we send three physical bits, all in state 0 and each time we want to send logical 1 we send three physical bits, all in state 1. The receiver decodes the bit value by a "majority vote" of the three bits. If only one error occurs then this error correction procedure is foolproof. In general the net probability of error is just the likelihood that two or three errors occur, which is $3p^2(1-p) + p^3 \leq p$. Thus, the three bit code improves the reliability of the information transfer. The quantum case is more complicated because we have both bit-flip and phase-flip errors.

1.5. **Quantum error correction.** In order to protect a qubit against bit-flips (incoherent X rotations), we rely on the same repetition code, but both encoding and error correction is now done by quantum operations. We take a qubit in some unknown pure state $\alpha|0\rangle + \beta|1\rangle$, bring two auxiliary qubits, and encode it into the

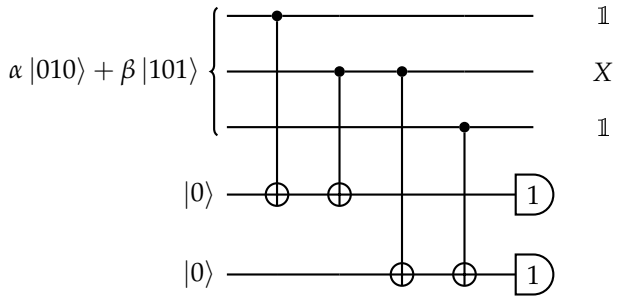
three qubits as



Suppose at most one qubit is then flipped, say the second one. The encoded state becomes

$$\alpha|010\rangle + \beta|101\rangle .$$

Decoding requires some care. Note that if we measured the three qubits directly, that would destroy the superposition of states which we are working so hard to protect. Instead we bring two additional qubits, both in state $|0\rangle$, and apply the following network



We measure the two auxiliary qubits, also known as ancillas, and the result of the measurement, known as the error syndrome, tells us how to reset the three qubits of the code. The theory behind this network runs as follows. If qubits one and two, counting from the top, are the same, then the first ancilla is in the $|0\rangle$ state. Similarly, if qubits two and three are the same, then the second ancilla is in the $|0\rangle$ state. However, if they are different, the corresponding ancilla is in the $|1\rangle$ state. Hence, the four possible error syndromes $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$ each indicate a different possibility – no errors, an error in the third, first or second qubits respectively. In our example, we would measure $|11\rangle$, revealing that both qubits 1 and 2, and qubits 2 and 3, are different. Thus, it is qubit 2 that has an error on it. Knowing the error, we can go back and fix it, simply by applying X to qubit 2. The net result is the state $\alpha|000\rangle + \beta|111\rangle$, which is then turned into $(\alpha|0\rangle + \beta|1\rangle)|0\rangle|0\rangle$ by running the mirror image of the encoding network.

1.6. **Turning bit-flips to phase-flips.** The three-qubit code that we have just demonstrated is sufficient to protect a qubit against single bit-flips, but not phase-flips. But this is good enough. Recall that $HZH = X$, hence it is enough to sandwich the decoherence area in between the Hadamard gates - they will turn phase flips into bit flips - and we can protect our qubits against Z-errors. The encoded state $\alpha|0\rangle + \beta|1\rangle$ now reads $\alpha|+++\rangle + \beta|---\rangle$, where $|\pm\rangle = |0\rangle \pm |1\rangle$.

1.7. **Dealing with bit-flip and phase-flip errors.** We can now put the bit-flip and phase-flip codes together: first we encode the qubit using the phase-flip code and then we encode each of the three qubits of the code using the bit-flip code. This gives an error correction scheme that allows us to protect against both types of error, thus yielding a code that encodes a single logical qubit across nine physical qubits, protecting against a single quantum error on any of the nine qubits.

If we want to preserve a quantum state for a long time without doing any computations, or if we want to send it through a noisy communications channel, we can just encode the state using a quantum code and decode it when we are done.

Computation on encoded states using noisy gates requires few more tricks (to be completed).