Supplementary Materials for Quantum Computing (CST Part II)

Steven Herbert

2020–21

Abstract

These notes contain supplementary material to complement the information in the lectures. Thus together, these notes and the lecture slides comprise a self-contained set of lecture notes. For consistency, the section numbering herein corresponds to the lecture numbering. These additional notes do not attempt to capture *everything* of interest which is tangential to the core course material, as this would simply be too large a scope, and so for further information the reader is pointed to the list of references on the course website.

The material in these notes is non-examinable, apart from where the material is a repeat of content also covered in the lectures (for example projective measurements, which are covered in Lecture 11 of the course, but are also summarised in these supplementary notes for Lecture 4).

Contents

1 Bits and Qubits

1.1 From double-slits to qubits

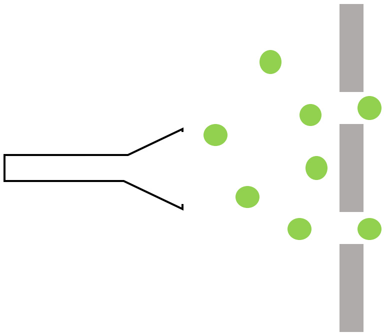
The double-slit experiment was one of the first demonstrations of quantum phenomena and so is an appropriate starting point when trying to gain a conceptual grasp of the essential, physical nature of quantum mechanics. It also provides a nice visualisation of why relative phase matters but global phase doesn't, and gives some insight into exactly what it means "to measure a quantum system".

1.1.1 Particles and waves

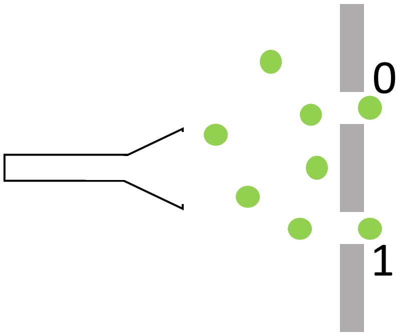
The double-slit experiment illustrates that sub-atomic particles demonstrate "wave-particle duality", and we start by spelling out exactly how particles and waves behave when passing through a screen with two slits cut in it.

Particles

Consider a machine firing out balls at a screen that has two slits in:

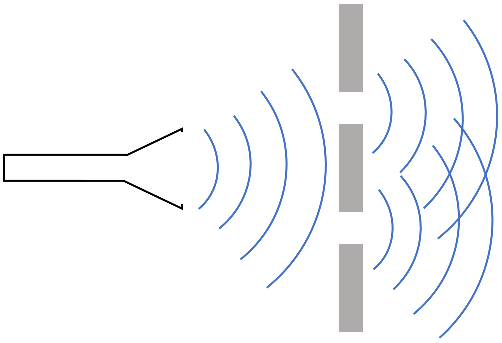


Any ball that passes through to the right-hand side of the screen must have passed through one of the slits. As there are two slits, there are two possibilities, so **we can record which slit a given ball has passed through with a single bit**. For example, we may use the value "0" to record that the ball has passed through the upper slit (as we view it on the page) and the value "1" to record that the ball has passed through the lower slit:



Waves

Now consider the situation in which a light source emits light (that is, an electromagnetic wave), which travels towards a screen with two slits:



The electric field can be used to describe the effect of a propagating electromagnetic wave at a given point in space, so we now consider the electric field at the transmitting source, which in general can be expressed (in a simplified manner) as:

$$E = E_0 e^{i\omega t} \tag{1}$$

where E is the electric field, E_0 is a constant, ω is the angular frequency ($\omega = 2\pi f$ where f is the frequency) and t is time. This oscillating electric field then propagates at the speed of light, so if the distance to each of the slits is d, then if we denote the locations of the upper and lower slits "$\mathbf{u}$" and "$\mathbf{l}$" respectively, we get that the electric field at the slits is:

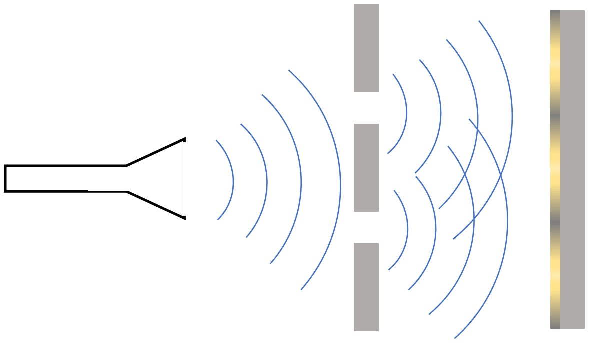
$$E(\mathbf{u}) = E_u e^{i(\omega(t-(d/c)))} = E_u e^{i\omega t} e^{-i\omega d/c} \tag{2}$$

for some constant E_u, and where c is the speed of light (note the constant E_u is needed as, in general, there will be a reduction in the magnitude of the electric field as the wave propagates and the radiation spreads out so we cannot simply reuse E_0 – but the specifics of this process do not concern us here). Thus the d/c term simply represents the time lag incurred by the wave travelling to the slit. We can also express the electric field at the lower slit

$$E(\mathbf{l}) = E_l e^{i\omega t} e^{-i\omega d/c} \tag{3}$$

for some constant E_l.

We now turn out attention to the light propagation beyond (to the right-hand side of) the screen. The electric field to the right-hand side corresponds to the superposition (sum) of electric fields $E(\mathbf{u})$ and $E(\mathbf{l})$ emanating from the upper- and lower-slits respectively. That is, the wave propagates as if there were two light sources, one at either slit. This means that waves from the two light sources (really, components of the single light-source corresponding to the two slits) will interfere. If we now place a second screen to the right-hand side of the existing double-slitted screen, we will observe an interference pattern, corresponding to the constructive and destructive interference:



1.1.2 Wave particle duality

Maxwell's laws formalise the behaviour of all electromagnetic phenomena (including light) in terms of waves, and these were very successful at accurately explaining most observed phenomena. However, some experiments suggested that, instead, light was actually composed of tiny particles, named "photons". The double-slit experiment (with a second screen to detect an interference pattern) was therefore a promising candidate to resolve this quandary, however the result was even more baffling. If the light was shone at the double-slitted screen, then an interference pattern was indeed observed on the second screen, indicating wave-light behaviour. However, if measurement apparatus was positioned at each of the double-slits to try to detect which of the two slits each photon passed through, then something amazing happened: the measurement apparatus did indeed detect photons passing through one slit or the other, but the interference pattern disappeared on the second screen.

This behaviour was observed even when the light source was dimmed sufficiently that only a single-photon was emitted at a time. With the measurement apparatus, each photon was observed to pass through one of the slits; but without the measurement apparatus an interference pattern was observed at the second screen, apparently showing that even a single photon can propagate as a wave, pass through both slits and subsequently interfere with itself.

This weird behaviour has become known as *wave-particle* duality, and can be thought of in the following way: unobserved a quantum system evolves as a wave (i.e., a wave passing through two slits and subsequently interfering with itself) – but when measured as if it has objective (classical) reality, the wave-function collapses and it does indeed have objective reality (i.e., as a photon or if you like a little "ball" passing through one of the two slits).

So it follows that our mathematical description of a quantum system should be sufficient to allow both possibilities – it should both enable the (probabilistic) measurement outcomes to be determined, and also fully capture the subsequent wave propagation (if a measurement isn't made). In particular, according to Postulate 1 of quantum mechanics (see Lecture 3), *the system is completely described by its state vector*, thus the quantum state at the double slit must completely capture everything about the wave-particle duality. For a two-level quantum system (qubit), we can qualitatively appreciate that a complex superposition over computational basis vectors has the required ingredients. The computational basis vectors ($|0\rangle$ and $|1\rangle$) represent the binary states which can occur if measured (i.e., which slit the photon has passed through) – and the complex coefficients thereof allow the probabilities of each to be evaluated, but also are sufficient to allow the subsequent wave-propagation (i.e., to the right-hand side of the screen) to be expressed if a measurement isn't made (and this is why they must be complex).

This also provides a nice way to think about computational bases, as in some sense representing "classical" events with objective reality, and measurement thereof as simply obtaining and ascertaining this classical reality by collapsing the wave-function. This also de-mystifies what, exactly, (computational basis) quantum measurement is, to some extent: is just normal measurement with a voltmeter, an ammeter or a signal analyser etc. General measurement (i.e., according to the measurement postulate in Lecture 3) *is* somewhat less tangible, but for the purposes of the Part II CST Quantum Computing Course, we almost invariably use computational basis measurement, which has this precise physical interpretation.

What about the philosophy of the measurement problem?

Speaking of quantum *measurement* implicitly relies on an objective distinction between a quantum system being measured, and a classical system doing the measurement. But physically, all this comes down to is that "small" objects are quantum, and "large" (i.e., not microscopic) objects are classical. The problem is that we know that these large classical objects are themselves composed of microscopic particles, and so there is a somewhat arbitrary, and certainly unsatisfactory distinction between the classical and quantum worlds, which is manifested when performing a measurement. This is known as the *measurement problem* in philosophy. In this course we take an operational approach and simply accept that this distinction is clear in practise, and thus quantum objects and classical measurement apparatus are well-defined, and we do not trouble ourselves with precisely how this distinction occurs. This approach is aligned with the *Copenhagen interpretation* of quantum mechanics – in fact, it would be fair to say that the Copenhagen interpretation has been embedded in the very way we've looked at quantum mechanics thus far – with concepts like classical measurement outcomes of quantum events being expressed using the Born rule being characteristic of the Copenhagen interpretation.

There are, however, alternative interpretations of quantum mechanics, which do seek to resolve the apparent measurement problem. To discuss some of these, it is helpful to introduce *Schrödinger's cat* thought experiment, for which the measurement problem serves as the premise. In the thought experiment, a cat is placed in a closed box, alongside a vial of deadly poison. A quantum system is set up (usually considered to be the decay of some sub-atomic particle), so that if the decay occurs then the vial will break, and the cat will be killed. However, as the particle's decay is described as by a quantum system, it will actually be in a superposition of decayed and not-decayed until it is measured. But what about the cat? Well, in quantum mechanical parlance, it has become entangled with the decaying particle, and hence it is in a superposition of being dead and alive. That is, until somebody opens the box, to observe (measure) whether it is dead or alive the cast is both dead and alive!

Schrödinger intended his thought experiment to be an absurdity with which the Copenhagen interpretation could be refuted, but with the debate still raging 85 years later, it has proven to be anything but. However, Schrödinger's cat does raise one subtlety about the nature of measurement itself – namely, is classical measurement just a proxy for conscious observation? The *consciousness causes collapse* interpretation holds that it is, and hence it is our very ability, as sentient "conscious" entities, to witness nature that leads to wavefunction collapse of quantum systems. Whether the cat is a conscious entity that it therefore capable of fulfilling its own demise by observation is an unresolved point.

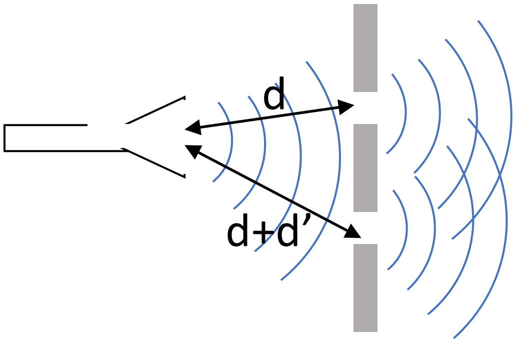
Other attempts to resolve the measurement problem include the *spontaneous collapse* interpretation, which holds that the linear, unitary, nature of quantum mechanics is merely a very, very good approximation – but in reality, it is only an approximation, and this is manifested by particles in superposition occasionally spontaneously collapsing. These occurrences of spontaneous collapse are proposed to be phenomenally rare for single particles moving freely (that is, evolving according to their wavefunction). But when particles are bonded together, a process akin to collapse is much more common, enabling classical objects to "emerge" from the underlying quantum world. So it follows that when quantum objects come into contact with classical objects (as is the process of measurement), a collapse is highly likely (that is, to all intents and purposes, certain) to occur.

For a variety of reasons, neither the *spontaneous collapse* interpretation nor the *consciousness causes collapse* interpretation have gained too much mainstream attention, however one resolution to the measurement problem has, namely the *many worlds* interpretation. In the *many worlds* interpretation of quantum mechanics, each time a measurement occurs there is a branching process. Say a certain measurement has two possible outcomes, then when the measurement takes places, the universe branches into two non-communicating parallel universes, one each for the two possible outcomes. The set of these parallel universes is usually termed the multiverse and viewing the multiverse as a whole, we can see that there is no actual wavefunction collapse – by treating the component universes of the multiverse as being in superposition, there is no significance to the moment of measurement. In this scenario wavefunction collapse is merely an illusion – an artefact of the fact that the observer themselves branches (into two non-communicating parallel copies) so that one copy observers each of the two outcomes.

Taking a step-back, these are *interpretations* of quantum mechanics precisely because they lie outside of quantum mechanics itself. As such, there is no overall consensus about the "true" interpretation. Personally, I normally go about my research by the same mantra that I teach: that the distinction between the classical and the quantum is clear in practise, and don't trouble myself further – although I am occasionally drawn to ponder the fundamental nature of reality in a more philosophical way.

1.1.3 Global and relative phase

The double-slit experiment also provides us with a sketch of why relative phase is important, but global phase isn't (Lecture 3).



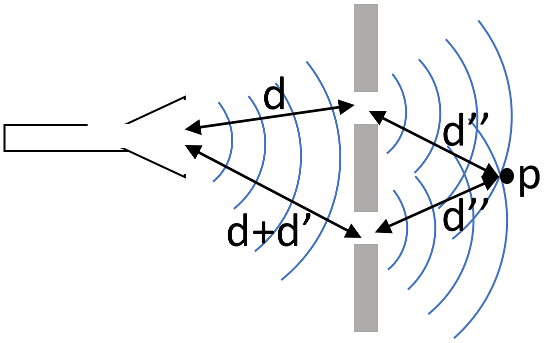
If we adjust the double slit so that the lower of the slits is now a distance $d + d'$ from the source, as shown above, we get electric fields at the upper slit

$$E(\mathbf{u}) = E_u e^{i\omega t} e^{-i\omega d/c} \tag{4}$$

as before, but for the lower slit

$$E(\mathbf{l}) = E_l e^{i\omega t} e^{-i\omega(d+d')/c} = E_l e^{i\omega t} e^{-i\omega d/c} e^{-i\omega d'/c} = E_l e^{i\omega t} e^{-i\omega d/c} e^{-i\phi} \tag{5}$$

where we define $\phi = \omega d'/c$ (which we can do because the angular frequency is a constant). If we now want to know the electric field at some point "$\mathbf{p}$" equidistant from the two slits (and to the right-hand side of by a distance d''),



we simply add the electric fields (i.e., because of linearity of electric field):

$$\begin{aligned} E(\mathbf{p}) =& E'_u e^{i\omega t} e^{-i\omega d/c} e^{-i\omega d''/c} + E'_l e^{i\omega t} e^{-i\omega d/c} e^{-i\phi} e^{-i\omega d''/c} \\ =& e^{-i\omega d/c} e^{-i\omega d''/c} e^{i\omega t} (E'_u + E'_l e^{-i\phi}) \end{aligned} \tag{6}$$

where the constants E'_u and E'_l have been defined to allow for further reduction in electric field magnitude owing to the further propagation. If we let $E'_u \approx E'_l$ then we can express this as:

$$E(\mathbf{p}) = E'_u e^{-i\omega(d+d'')/c} e^{i\omega t} (1 + e^{-i\phi}) \tag{7}$$

We can see that $E'_u e^{-i\omega(d+d'')/c}$ is a constant that has been “factored out”, and the constant $-\omega(d+d'')/c$ is the *global phase*, which has only a classical effect. However, the quantum element of the wave’s behaviour only concerns how the two superposed components interfere (and thus the probabilities of measurement outcomes at various points in the evolution), and this is determined only by the *relative phase* $-\phi$. For example, if the difference in path length between the two components, d', is such that $\phi = (2n+1)\pi$, then the interference will be destructive, and if its such that $\phi = 2\pi$ it will be constructive (for some integer $n \geq 0$; i.e., this is because $e^{-i2n\pi} = 1$, $e^{-i(2n+1)\pi} = -1$). This is true regardless of the global phase.

In the Part II CST Quantum Computing Course we deal exclusively with *pure quantum states*, however in quantum information theory it is common to analyse systems in terms of classical mixtures of quantum states, or *mixed quantum states* as they are known. In this case, there is a classical element in the analysis and so the global phase may be relevant (although frequently the global phase can still be neglected even in the scenario).

1.1.4 Limitations of the analogy

The wave-particle duality exhibited in the double-slit experiment is useful for illustrating why qubits (and indeed general quantum systems) are represented as they are – as complex superpositions over the computational basis states, but in other ways the analogy is limited. In particular, the propagation of a photon is a complicated quantum event, and it is only by viewing it at the double-slit itself that we can (in a slightly contrived way) think of it as a two-level quantum system. Whilst the quantum state could be normalised such that at the double-slit it indeed appears as a 2-element complex unit vector, in reality the system is more complicated than this, and so it follows that there is no real interpretation within the double-slit / qubit analogy for the unitary evolution that an actual qubit can undergo. Furthermore, a crucial aspect of quantum information processing and computing is how qubits interact (including entangling interactions), but again the double-slit experiment cannot readily be used to explain this.

There exist in nature more genuine two-level quantum systems, that do indeed evolve unitarily and can interact with one another. For students who are interested, the *Stern-Gerlach experiment* provides another example of a two-level quantum system, which allows better explanation of some of these further aspects of the behaviour and nature of qubits.

3 The Postulates of Quantum Mechanics

3.1 The Bloch sphere

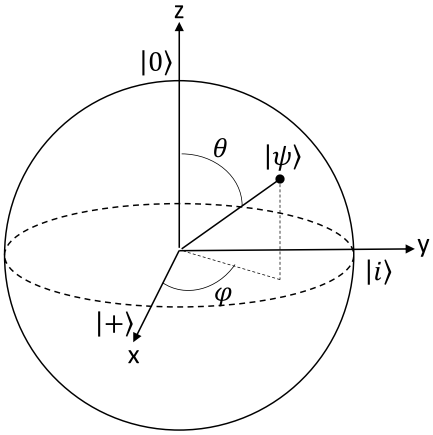
In the lecture we saw that any single-qubit quantum state can be expressed as:

$$|\psi\rangle = e^{i\omega}(\alpha\,|0\rangle + \beta e^{i\varphi}\,|1\rangle) \tag{8}$$

where α and β are real and positive (note ω is used in place of θ, which was used in the lecture, to avoid ambiguity with what follows). Owing to the fact that we can ignore the global phase and that $|\alpha|^2 + |\beta|^2 = 1$, we can express (8) as:

$$|\psi\rangle = \cos(\theta/2)\,|0\rangle + e^{i\varphi}\sin(\theta/2)\,|1\rangle \tag{9}$$

for some angles θ and φ. This in turn leads to a nice illustration of a single qubit state, as any point on the surface of the *Bloch sphere* (strictly, any single qubit *pure* state – however we do not deal with mixed quantum states in this course):



A few important notes about the Bloch sphere:

- Orthogonal states are co-polar points on the Bloch sphere surface (that is, they are not at an angle of 90° from each other, as one may intuitively expect). It is always possible to perform a measurement that perfectly distinguishes states that are co-polar in the Bloch sphere (i.e., orthogonal states – see Lecture 4), but this is not true for any pair of non co-polar points.

- Three particularly noteworthy pairs of orthogonal states are those that lie along the x, y and z axes: these are $|0\rangle, |1\rangle$; $|i\rangle, |{-i}\rangle$; and $|+\rangle, |-\rangle$ respectively (one of each pair is shown in the figure, the other is simply at point "-1" along the corresponding axis). We have met $|0\rangle, |1\rangle$ and $|+\rangle, |-\rangle$ in the lecture, and also have that $|i\rangle = (1/\sqrt{2})(|0\rangle + i\,|1\rangle)$ and $|{-i}\rangle = (1/\sqrt{2})(|0\rangle - i\,|1\rangle)$.

- As a single qubit unitary maps a single-qubit state to another single-qubit state; we can see that a unitary transformation is itself an operation that maps all of the points on the surface of the Bloch to other points on the surface of the Bloch sphere (in general, different points on the Bloch sphere – the exception being when the unitary is the identity).

- The Pauli X, Y and Z operations are rotations of π radians around the x, y and z axes respectively.

- The Bloch sphere can also be used to illustrate three other *parameterised* quantum gates: the rotation gates $R_x(\sigma)$, $R_y(\sigma)$ and $R_z(\sigma)$, which are unitaries that rotate the state by σ radians around the x, y and z axes, respectively.

4 Important Concepts in Quantum Mechanics

In the lecture we touch on many important concepts in quantum mechanics – for some of these, in the lecture itself we prioritise gaining an overall understanding of the concept itself, over a rigorous proof. Therefore in these supplementary materials we include some more rigorous additional results, as well as other tangential material of background interest. One omission is a proof of the Helstrom-Holevo bound, for this requires significant pre-requisites in quantum information theory, which are beyond the scope of even these supplementary materials. However, the (very) keen student is pointed to Nielsen and Chuang for a full explanation and derivation of this.

4.1 Multi-qubit measurement

Quantum computing is often described as the child of physics and computer science – and it is perhaps the case that the aspect of the field that these parents are most in conflict over is the subject of quantum measurement. To think of measurement as ascertaining some quantity pertaining to a physical system as a binary string is second nature to computer scientists; but physicists have a more nebulous idea of what constitutes *measurement*, and this in turn has informed the presentation of the measurement postulate. We will now show that the general measurement postulate, complemented with the ability to perform unitary operations, ultimately amounts to the same thing that we consider to be measurement in computer science: that is, measurement in the computational basis.

In the lecture we saw that any single-qubit orthogonal-basis measurement can be achieved by a unitary transformation followed by a computational basis measurement. This is important, as we previously argued that computational basis measurements have some tangible physical interpretation of extracting classical information from a quantum state (i.e., via an ordinary, albeit precise, measuring apparatus), and thus collapsing the quantum state accordingly.

Equally importantly, though, is the fact that we can achieve any n-qubit orthogonal-basis measurement by n-qubit unitary transformation, followed by computational basis measurements on each of the n qubits. By an "n-qubit orthogonal-basis measurement", we mean a measurement where the measurement operators (as defined in Lecture 3) are projectors onto the orthogonal basis vectors. Such measurement is also termed *projective measurement*. Let $|1\rangle \dots |N\rangle$ be the $N = 2^n$ computational basis vectors (recall that the i^{th} computational basis vector is a vector of all zeros except for a single 1 in the i^{th} element), and let $|b_1\rangle \dots |b_N\rangle$ be a general orthonormal basis. Next we define the projectors on the orthonormal basis:

$$M_i = |b_i\rangle \langle b_i| \tag{10}$$

for all $i = 1 \dots N$. We also define the unitary (slightly abusing notation by using states as column vectors that are then concatenated to form a matrix):

$$U = \begin{bmatrix} |b_1\rangle & |b_2\rangle & \dots & |b_N\rangle \end{bmatrix} \tag{11}$$

which is thus such that:

$$|b_i\rangle = U|i\rangle \tag{12}$$

So we can see that the probability of the i^{th} measurement outcome when performing a projective measurement onto the orthonormal basis $|b_1\rangle \dots |b_N\rangle$ is (starting with the definition from Lecture 3):

$$\begin{aligned} p_i &= \langle\psi| M_i^\dagger M_i |\psi\rangle \\ &= \langle\psi| |b_i\rangle \langle b_i| |b_i\rangle \langle b_i| |\psi\rangle \\ &= \langle\psi| U |i\rangle \langle i| U^\dagger U |i\rangle \langle i| U^\dagger |\psi\rangle \\ &= ((\langle\psi| U)(|i\rangle \langle i|)(|i\rangle \langle i|)(U^\dagger |\psi\rangle)) \end{aligned} \tag{13}$$

which we can thus see amounts to first applying the unitary transform $U^\dagger$ to the state being measured, and then measuring in the computational basis – i.e., where $|i\rangle\langle i| = (|i\rangle\langle i|)^\dagger$ are the measurement operators. After the measurement the state will be the unitarily transformed state, collapsed into the state corresponding to the measurement outcome.

In quantum mechanics projective measurements are usually used in the context of measuring observables. An observable is a Hermitian operator, say H, which has spectral decomposition:

$$\mathrm{H} = \sum_i i\,|e_i\rangle\langle e_i|\,, \tag{14}$$

where i is the eigenvalue corresponding to eigenvector $|e_i\rangle$ and we let $P_i = |e_i\rangle\langle e_i|$ denote the projector onto the i^{th} eigenspace. As projectors are always such that $P_i^\dagger P_i = P_i$, we get that the probability of measuring the i^{th} eigenvector as (when measuring some arbitrary state $|\psi\rangle$):

$$p(i) = \langle\psi| P_i |\psi\rangle\,. \tag{15}$$

Crucially, if we get the i^{th} eigenvector from the measurement, then we *interpret the measurement outcome as obtaining a numerical value equal to the i^{th} eigenvalue.* This leads to the notion of measuring the expectation of an observable:

$$\begin{aligned} \mathbb{E}(\mathrm{H}) &= \sum_i i\,p(i) \\ &= \sum_i i\,\langle\psi| P_i |\psi\rangle \\ &= \langle\psi| \left(\sum_i iP_i\right) |\psi\rangle \\ &= \langle\psi| \mathrm{H} |\psi\rangle \end{aligned} \tag{16}$$

4.2 More on the no-signalling principle

In the lecture we looked at a simple instance of the no-signalling principle, in which Alice and Bob share a Bell pair of which they then each measure one qubit. We saw that Bob's measurement statistics are unaffected by whether or not Alice has yet measured her qubit. We now show that the same holds for a slightly more general case in which:

- Alice and Bob share a general two-qubit entangled state.
- Alice applies a unitary operation to her qubit and then measures.
- Bob applies a unitary operation to his qubit and then measures.

In particular, we will show that Bob measures zero with the same probability regardless of whether Alice has already measured her qubit or not (note that we consider computational basis measurement, but we can see that by allowing Alice and Bob to perform a unitary operation on their qubits prior to measurement, we implicitly consider general single-qubit measurements).

To see this, let the (entangled) state that Alice and Bob share be:

$$\alpha |00\rangle + \beta |01\rangle + \gamma |10\rangle + \delta |11\rangle \tag{17}$$

where Alice holds the first qubit. We also let Alice's unitary operation be $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, so the action on the entire state is $A \otimes I$ and Bob's unitary operation be $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$, so the action on the entire state is $I \otimes B$. In the case where Alice performs her measurement first, we get the pre-measurement state:

$$(\alpha A_{11} + \gamma A_{12}) |00\rangle + (\beta A_{11} + \delta A_{12}) |01\rangle + (\alpha A_{21} + \gamma A_{22}) |10\rangle + (\beta A_{21} + \delta A_{2}) |11\rangle \tag{18}$$

Which means that if Alice measures $|0\rangle$, the state collapses to

$$\frac{1}{\sqrt{|\alpha A_{11} + \gamma A_{12}|^2 + |\beta A_{11} + \delta A_{12}|^2}} |0\rangle \otimes ((\alpha A_{11} + \gamma A_{12}) |0\rangle + (\beta A_{11} + \delta A_{12}) |1\rangle)) \tag{19}$$

which occurs with probability $|\alpha A_{11} + \gamma A_{12}|^2 + |\beta A_{11} + \delta A_{12}|^2$. If Alice measures $|1\rangle$ the state collapses to:

$$\frac{1}{\sqrt{|\alpha A_{21} + \gamma A_{22}|^2 + |\beta A_{21} + \delta A_{22}|^2}} |1\rangle \otimes ((\alpha A_{21} + \gamma A_{22}) |0\rangle + (\beta A_{21} + \delta A_{22}) |1\rangle)) \tag{20}$$

with probability $|\alpha A_{21} + \gamma A_{22}|^2 + |\beta A_{21} + \delta A_{22}|^2$. Taking these two cases in turn, in the first (when Alice measures 0) Bob get measurement outcome 0 with probability:

$$\left(\frac{1}{\sqrt{|\alpha A_{11} + \gamma A_{12}|^2 + |\beta A_{11} + \delta A_{12}|^2}} \right)^2 |B_{11}(\alpha A_{11} + \gamma A_{12}) + B_{12}(\beta A_{11} + \delta A_{12})|^2 \tag{21}$$

In the case where Alice has measured one, Bob measures 0 with probability:

$$\left(\frac{1}{\sqrt{|\alpha A_{21} + \gamma A_{22}|^2 + |\beta A_{21} + \delta A_{22}|^2}} \right)^2 |B_{11}(\alpha A_{21} + \gamma A_{22}) + B_{12}(\beta A_{21} + \delta A_{22})|^2 \tag{22}$$

Noting that the first case occurs with probability $|\alpha A_{11} + \gamma A_{12}|^2 + |\beta A_{11} + \delta A_{12}|^2$ and the second with probability $|\alpha A_{21} + \gamma A_{22}|^2 + |\beta A_{21} + \delta A_{22}|^2$ we get the overall probability of Bob measuring a zero as:

$$P_B(0) = |B_{11}(\alpha A_{11} + \gamma A_{12}) + B_{12}(\beta A_{11} + \delta A_{12})|^2 + |B_{11}(\alpha A_{21} + \gamma A_{22}) + B_{12}(\beta A_{21} + \delta A_{22})|^2 \tag{23}$$

Which we can manipulate as follows:

$$\begin{aligned}
P_B(0) &= \left| \begin{bmatrix} B_{11} & B_{12} \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \begin{bmatrix} A_{11} \\ A_{12} \end{bmatrix} \right|^2 + \left| \begin{bmatrix} B_{11} & B_{12} \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \begin{bmatrix} A_{21} \\ A_{22} \end{bmatrix} \right|^2 \\
&= \begin{bmatrix} B_{11} & B_{12} \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \begin{bmatrix} A_{11} \\ A_{12} \end{bmatrix} \begin{bmatrix} A_{11}^* & A_{12}^* \end{bmatrix} \begin{bmatrix} \alpha^* & \beta^* \\ \gamma^* & \delta^* \end{bmatrix} \begin{bmatrix} B_{11}^* \\ B_{12}^* \end{bmatrix} \\
&\quad + \begin{bmatrix} B_{11} & B_{12} \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \begin{bmatrix} A_{21} \\ A_{22} \end{bmatrix} \begin{bmatrix} A_{21}^* & A_{22}^* \end{bmatrix} \begin{bmatrix} \alpha^* & \beta^* \\ \gamma^* & \delta^* \end{bmatrix} \begin{bmatrix} B_{11}^* \\ B_{12}^* \end{bmatrix} \\
&= \begin{bmatrix} B_{11} & B_{12} \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \begin{bmatrix} |A_{11}|^2 + |A_{21}|^2 & A_{11}A_{12}^* + A_{21}A_{22}^* \\ A_{11}^*A_{12} + A_{21}^*A_{22} & |A_{12}|^2 + |A_{22}|^2 \end{bmatrix} \begin{bmatrix} \alpha^* & \beta^* \\ \gamma^* & \delta^* \end{bmatrix} \begin{bmatrix} B_{11}^* \\ B_{12}^* \end{bmatrix} \qquad (24)
\end{aligned}$$

Because A is unitary its columns are orthonormal vectors, and we can see that this means that the middle matrix is equal to the identity. We can also multiply out the remaining to see that:

$$P_B(0) = |\alpha B_{11} + \beta B_{12}|^2 + |\gamma B_{11} + \delta B_{12}|^2 \qquad (25)$$

Turning now to case in which Bob measures his qubit prior to Alice measuring hers we see that his unitary transforms the original state to:

$$(\alpha B_{11} + \beta B_{12})\,|00\rangle + (\alpha B_{21} + \beta B_{22})\,|01\rangle + (\gamma B_{11} + \delta B_{12})\,|10\rangle + (\gamma B_{21} + \delta B_{22})\,|11\rangle \qquad (26)$$

from which we can see that we also get the probability of Bob measuring 0 as:

$$|\alpha B_{11} + \beta B_{12}|^2 + |\gamma B_{11} + \delta B_{12}|^2 \qquad (27)$$

i.e., the same probability as for the case where Alice measured first. So we can see that Bob cannot detect whether of not Alice has performed her measurement by measuring his qubit, and so the no-signalling principle is upheld.

Of course it is possible to go even more general still, and consider any case in which Alice and Bob are spatially separated and share any kind of entangled resource (potentially with third parties involved as well), however even then the no-signalling principle holds – and there is little additional insight to be gained by exhaustively working through all of the possibilities.

4.3 Bell's theorem

The no-signalling principle shows that entanglement cannot be used for super-luminal (also termed "non-local") signalling, however entanglement can be shown to lead to observable non-local physical events.

The story begins with a trio of august physicists, Einstein, Podolsky and Rosen (EPR), who were adamant that entanglement must in fact merely amount to the presence of a hidden variable, and that the apparent entangled superpositions were therefore merely a (classical) lack of information about the nature of the hidden variable. For example, if we consider the entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, then if we only ever measure in the computational basis, we actually have no way of knowing that we are measuring an entangled state rather than a *classical mixture* of the (classical) states $|00\rangle$ and $|11\rangle$. In particular, in this case we would still always get the same measurement outcome from each qubit (0 or 1), and each with probability one half. However, rather than this being due to the entangled superposition, it would merely be down to a lack of information about a "hidden variable" whose identity pre-determines the measurement outcomes. Enter the stage John Stewart Bell who realised that the secret to distinguishing a *classical mixture* (any sample from which could be thought of as being pre-determined by a hidden-variable) from an *entangled quantum state* whose measurement outcomes are not pre-determined is to allow measurement in different bases. This led him to propose a remarkable well-defined physical experiment to detect the presence of entanglement.

To do this, he began by supposing that two spatially separated parties, by convention "Alice" and "Bob", are each issued by "Charlie" one half of a physical system, which they can thereafter perform a measurement on. Additionally, Charlie can prepare multiple copies of the same physical system to distribute to Alice and Bob. In the version of Bell's theorem we shall study, Alice and Bob are each capable of performing two different measurements on their system, and each of the four measurements (two each for Alice and Bob) can yield outcomes ±1.

Formally, let Alice's system have two properties A_1 and A_2 (each of which takes value ± 1 as specified above); and likewise let Bob's system have two properties B_1 and B_2 (again, each of which takes value ± 1). Alice and Bob simultaneously choose which property of their half of the system to measure (i.e., Alice A_1 or A_2, Bob B_1 or B_2). For simplicity let them choose each with probability one half. We now consider the value of:

$$A_1B_1 + A_2B_1 + A_2B_2 - A_1B_2 = (A_1 + A_2)B_1 + (A_2 - A_1)B_2 \tag{28}$$

Clearly as A_1 and A_2 are ± 1 then exactly one of $(A_1 + A_2)$ and $(A_2 - A_1)$ is equal to zero and the other equal to ± 2; so as B_1 and B_2 are each ± 1 we can see that the value is ± 2. We have stated that it must be possible to prepare identical copies of the physical system, however we do not require that the properties being measured are deterministic. Instead, let the system be defined by a probability distribution (the hidden variables posited by EPR are samples from this distribution), which we define with the probability mass function:

$$p(A_1 = a_1, A_2 = a_2, B_1 = b_1, B_2 = b_2), \tag{29}$$

for all permutations of a_1, a_2, b_1, $b_2 = \pm 1$ (hereafter we shorten this to $p(a_1, a_2, b_1, b_2)$). Next we consider the expectation of the term in (28):

$$\begin{aligned}\mathbb{E}(A_1B_1 + A_2B_1 + A_2B_2 - A_1B_2) &= \sum_{a_1,a_2,b_1,b_2} p(a_1, a_2, b_1, b_2)(a_1b_1 + a_2b_1 + a_2b_2 - a_1b_2) \\ &\leq \sum_{a_1,a_2,b_1,b_2} p(a_1, a_2, b_1, b_2) \times 2 \\ &= 2 \end{aligned} \tag{30}$$

By linearity of expectation we also have that:

$$\mathbb{E}(A_1B_1 + A_2B_1 + A_2B_2 - A_1B_2) = \mathbb{E}(A_1B_1) + \mathbb{E}(A_2B_1) + \mathbb{E}(A_2B_2) - \mathbb{E}(A_1B_2), \tag{31}$$

which gives the term known as the Bell inequality:

$$\mathbb{E}(A_1B_1) + \mathbb{E}(A_2B_1) + \mathbb{E}(A_2B_2) - \mathbb{E}(A_1B_2) \leq 2 \tag{32}$$

Now we consider that the physical system Alice and Bob share is the Bell state $|\Psi^-\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right)$, and that the measurements they perform are measurements of observables. In particular:

$$A_1 = Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{33}$$

$$A_2 = X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{34}$$

$$B_1 = \frac{1}{\sqrt{2}}(-Z - X) = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \tag{35}$$

$$B_2 = \frac{1}{\sqrt{2}}(Z - X) = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} \tag{36}$$

It is easy to verify that each of these four observables have one eigenvalue equal to 1 and one equal to -1, so these do indeed correspond to measurements of the system that can return ± 1, as originally specified. As the two (simultaneous) measurements that Alice and Bob make together act on the whole quantum system, it is convenient to write the measurement as a projective measurement with a single observable. For instance, if Alice chooses to measure observable A_1 and Bob chooses to measure observable B_2 then together the measurement is a projective measurement with observable:

$$A_1 \otimes B_2 = Z \otimes \left(\frac{1}{\sqrt{2}}(Z - X)\right) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{37}$$

And similarly for $A_1 \otimes B_1$, $A_2 \otimes B_1$ and $A_2 \otimes B_2$.

At this juncture, it is important to clarify a few things. Firstly, the fact that Alice and Bob independently choose which observable to measure is not overly significant: we treat this choice as an objective (classical) fact, and so for any copy of the prepared physical system exactly one of $A_1 \otimes B_1$, $A_1 \otimes B_2$, $A_2 \otimes B_1$ and $A_2 \otimes B_2$ will have been measured. Nor is it particularly pertinent with which probability Alice and Bob choose each observable – just that each is chosen sometimes (and independently), so each of $A_1 \otimes B_1$, $A_1 \otimes B_2$, $A_2 \otimes B_1$ and $A_2 \otimes B_2$ will arise in the course of many repeats of the experiment. Additionally, it is not actually crucial that Alice and Bob perform the measurements *exactly* simultaneously: formally, it is necessary that they do so in sufficiently quick succession to preclude the possibility of signalling (that is, the time lapse between the measurements must be less than the time its would take for a ray of light to traverse the intervening distance). However, even if the measurements aren't exactly simultaneous we can still treat them in terms of the measurement of an observable applied to the entire system. Were we to instead treat the measurements as sequential then we would need to do a first partial measurement followed by the second measurement which is fiddly and amounts to the same thing anyway. Finally, it is important not to read too much into the fact that we *do* treat Alice and Bob's measurement as a single observable – as this is represented by a separable (can be decomposed as a tensor product) matrix this is purely a convenient way to perform the analysis and does not imply any entanglement or correlation in the measurement itself.

The benefit of treating the measurements of Alice and Bob as a single observable is that we can easily evaluate the expectation of Alice's measurement outcome multiplied by Bob's. That is because $\mathbb{E}(A_iB_j) = \langle\psi|\left(A_i \otimes B_j\right)|\psi\rangle$ for a general state $|\psi\rangle$ being measured (and for $i = \{0, 1\}$, $j = \{0, 1\}$). To see this, consider that if $M = M_1 \otimes M_2$ and M_1 has eigenvectors $|m\rangle$ with associated eigenvalues m; and M_2 has eigenvectors $|n\rangle$ with associated eigenvalues n then clearly $|m\rangle \otimes |n\rangle$

are the eigenvectors of M, with associated eigenvalues mn. Thus, by measuring the expectation of the observable corresponding to Alice *and* Bob's measurements (together), we automatically obtain $\mathbb{E}(A_i B_j)$ by measuring the observable $A_i \otimes B_j$. So we can proceed to evaluate the expectation for each of the four possible measurements on the state $|\Psi^-\rangle$. Firstly,

$$\mathbb{E}(A_1 B_2) = \langle\Psi^-| \, (A_1 \otimes B_2) \, |\Psi^-\rangle = \frac{1}{2\sqrt{2}} \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} = -\frac{1}{\sqrt{2}} \tag{38}$$

And we also get:

$$\mathbb{E}(A_1 B_1) = \langle\Psi^-| \, (A_1 \otimes B_2) \, |\Psi^-\rangle = \frac{1}{\sqrt{2}} \tag{39}$$

$$\mathbb{E}(A_2 B_1) = \langle\Psi^-| \, (A_2 \otimes B_1) \, |\Psi^-\rangle = \frac{1}{\sqrt{2}} \tag{40}$$

$$\mathbb{E}(A_2 B_2) = \langle\Psi^-| \, (A_2 \otimes B_2) \, |\Psi^-\rangle = \frac{1}{\sqrt{2}} \tag{41}$$

Together this gives us:

$$\mathbb{E}(A_1 B_1) + \mathbb{E}(A_2 B_1) + \mathbb{E}(A_2 B_2) - \mathbb{E}(A_1 B_2) = 2\sqrt{2} \tag{42}$$

which is a violation of the Bell inequality (32).

It has been shown with overwhelming experimental confidence that it is possible to prepare (quantum) physical systems that *do* lead to measurement outcomes that violate Bell's inequality. From this we can conclude that some assumption leading to (32) does not hold in the case of quantum systems. In fact we have already identified that it is the implicit assumption that the measurement outcomes correspond to hidden variables sampled from $p(a_1, a_2, b_1, b_2)$ that does not apply to the quantum setting. The significance of Bell's theorem cannot be over-stated: for the inequalities have provided emphatic experimental evidence of the non-classical behaviour of quantum systems. Bell's theorem is usually taken as a refutation of *local realism.* Realism means that a measurable quantity of a physical system has an objective value that is independent of that measurement actually being made – superposed quantum states do not exhibit realism. Locality means that a physical system can only affect another physical system that is within its *light-cone* (that is, at such a time and distance away that a ray of light, travelling at the speed of light, could have reached the latter) – the fact that the state of a spatially separated entangled pair *instantaneously* collapses violates this.

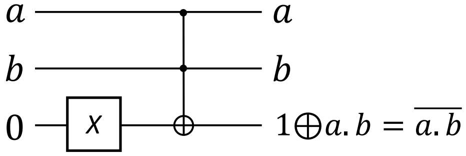
Bell's theorem kick-started the field of *quantum foundations* which seeks to understand the precise physical nature of the quantum world, but from a practical point of view the experimental violations of Bell's inequalities can be taken as evidence upholding the quantum mechanical postulates that we use to describe our quantum computing and information processing devices. Indeed, testing Bell inequalities is a frequently used way to investigate whether such a device really is exhibiting quantum behaviour.

5 The Quantum Circuit Model

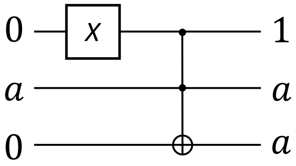
In the lecture we touched on two important notions of completeness of the quantum circuit model. Firstly, that it is sufficient to represent *any* unitary evolution; and secondly that any computation that can be performed classically can be performed on a quantum computer: and that the quantum circuit model is therefore sufficient to express this computation. Here we give further details of these, along with some other useful properties of quantum circuits.

5.1 Universal classical computation on quantum computers

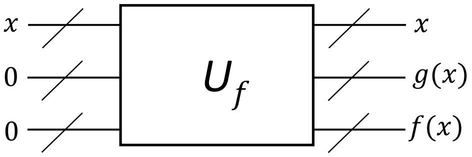
From classical logic, we know that together NAND and FANOUT constitute a universal gate-set. Each of these can be implemented using Toffoli and X gates:



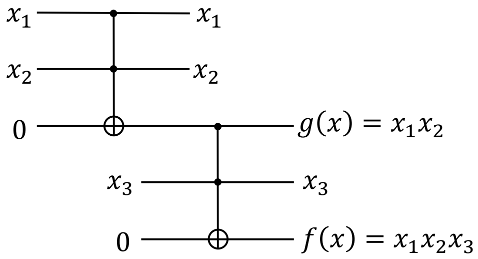
Note that an alternative would be to prepare the third qubit in the 1 state directly and thus allow us to omit the X gate, however for full generality we suppose that all ancilla qubits are prepared in the 0 state. Similarly, the FANOUT gate can be implemented:



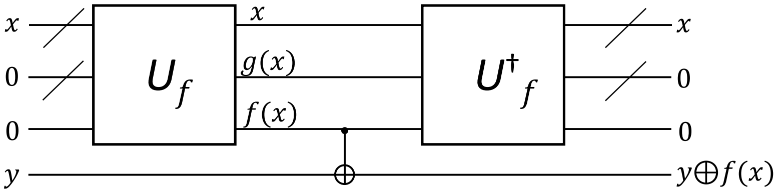
This means that any function that can be computed on a classical digital computer can be computed on a quantum computer, if a supply of ancillas is available as an auxilary input (for example, to prepare the 1 state required when the Toffoli acts as a FANOUT). This means, that we can compute any classically computable function, $f(x)$, on a quantum computer in the following way:



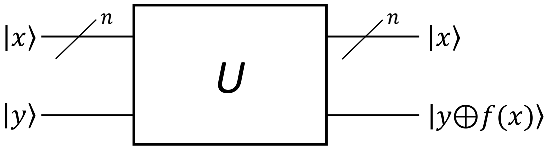
Where the block "U_f" is a quantum circuit consisting only of Toffoli and X gates and takes x as an input register (i.e., a binary string) and also an ancillary register (middle wire) and finally a register (of appropriate size) into which to put the result of the computation. The output $g(x)$ is the transformed state of the ancillary register, which will contain some intermediate results in the computation. For example consider a simple function which is the AND of three inputs, $f(x) = x_1.x_2.x_3$. This can be implemented using two Tofolli gates, but we can see that we get the additional output of $g(x) = x_1.x_2$.



Classically, the presence of $g(x)$ is not a problem, but quantumly it is: because our analysis generally assumes that there are no stray qubits entangled with the states used in the computation. To get around this, we use the fact that U_f is unitary and so has an inverse, $U_f^\dagger$, to perform a trick known as "uncomputing":



For simplicity, this has been shown for a function $f : \{0,1\}^n \to \{0,1\}$, as this enables us to use a single CNOT gate, however the same trick can be applied for any size of output register. When drawing quantum circuits, the ancilla register is typically omitted, so the *reversible form* of $f(x)$ is thus represented by the circuit block (now showing the inputs and outputs as quantum states, rather than just binary strings to emphasise that this is a circuit block):



As an aside, it is worth noting that the required number of ancillas used in the reversible circuit is at most proportional to the number of gates, which is always acceptable in principle (i.e., for complexity analysis), even if potentially problematic in practise.

5.2 More on universal gate-sets

As above, Toffolli and Pauli-X gates are together universal for classical computation, and therefore in principle we could use nothing but Toffoli and Pauli-X gates to compute any computable function on a quantum computer. That is because quantum computing has been shown not to violate the Church-Turing thesis, so there is nothing that can be computed on a quantum computer that cannot, in principle, be computed on a classical computer. Thus, as highlighted in the lecture, the belief is that the advantage of quantum computing is that it will enable the computation of some functions to be done (potentially exponentially) more efficiently. Clearly Toffoli and Pauli-X gates alone will not be sufficient to do this, so for quantum computing we have a more general notion of universality: (again as highlighted in the lecture) a universal gate-set is such that any unitary can be constructed with gates from the set. Consideration of universal gate-sets is important for both practical and theoretical reasons:

- In practise, any quantum computer will only be able to execute a restricted range of operations. It is therefore important to assert that even with such a restricted range, any overall unitary operation can be realised.
- In the theoretical analysis of quantum circuits we will often use gate-complexity interchangeably with time-complexity. However, this is only fair if we consider that all algorithms must be built from a finite set of gates. If, conversely, we allowed arbitrary unitaries to be considered as gates, then we could say that every quantum algorithm uses exactly one operation (i.e., whatever unitary the circuit represents), which is not a particular insightful result.

We will now show that CNOT, H, T is a universal gate-set, by first showing that any unitary can be decomposed into a circuit with only CNOTs and single-qubit unitaries; and then that any single-qubit unitary can be constructed from H and T gates. To do this, we first need to consider a couple of preliminary results concerning the rotation gates and controlled unitaries.

Rotation gates

Firstly, note that for any matrix, A, which is such that $A^2 = I$, we have that:

$$\exp(iAx) = \cos(x)I + i\sin(x)A \tag{43}$$

This allows us to introduce the following:

$$R_x(\theta) = \exp(-i\theta X/2) = \cos(\theta/2)I - i\sin(\theta/2)X = \begin{bmatrix} \cos\theta/2 & -i\sin\theta/2 \\ -i\sin\theta/2 & \cos\theta/2 \end{bmatrix} \tag{44}$$

$$R_y(\theta) = \exp(-i\theta Y/2) = \cos(\theta/2)I - i\sin(\theta/2)Y = \begin{bmatrix} \cos\theta/2 & -\sin\theta/2 \\ \sin\theta/2 & \cos\theta/2 \end{bmatrix} \tag{45}$$

$$R_z(\theta) = \exp(-i\theta Z/2) = \cos(\theta/2)I - i\sin(\theta/2)Z = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix} \tag{46}$$

where X, Y and Z are the Pauli matrices. These are known as the *rotation gates*, as they are unitaries that have the effect of rotating any single-qubit state by the amount θ around the X, Y and Z axes respectively, when represented using the Bloch sphere. An important decomposition of an arbitrary single-qubit unitary, U is:

$$U = e^{ia}R_z(b)R_y(c)R_z(d), \tag{47}$$

for some real numbers a, b, c, d. This can be seen by noting that U can be expressed:

$$U = e^{ia}\begin{bmatrix} e^{-i(b+d)/2}\cos(c/2) & -e^{-i(b-d)/2}\sin(c/2) \\ e^{i(b-d)/2}\sin(c/2) & e^{i(b+d)/2}\cos(c/2) \end{bmatrix} \tag{48}$$

and noting that real numbers a, b, c, d can be chosen such that U corresponds to any desired single-qubit unitary. Note that we keep the term corresponding to the global phase, e^{ia}, because we later consider *controlled* unitaries of this form – where global phase does matter.

There is, however, no reason why we must stick to rotations around the Pauli axes. Consider an arbitrary complex unit vector, $\hat{n} = (n_x, n_y, n_z)$, decomposed into components in the directions of the x, y and z axes of the Bloch sphere. We have that a rotation of any (single-qubit) unit vector, by an angle θ around $\hat{n}$ is achieved by the following unitary operation:

$$R_{\hat{n}}(\theta) = \exp(-i\theta(n_xX + n_yY + n_zZ)/2) = \cos(\theta/2)I - i\sin(\theta/2)(n_xX + n_yY + n_zZ) \tag{49}$$

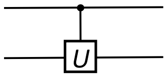
The proof of which is left to the reader. We also have that any single-qubit unitary can be decomposed into three rotations around any two fixed non-parallel axes $\hat{n}$ and $\hat{m}$, that is for any U we can find a', b', c', d' such that

$$U = e^{ia'} R_{\hat{n}}(b') R_{\hat{m}}(c') R_{\hat{n}}(d') \tag{50}$$

Again, the keen and/or skeptical reader is invited to prove this for themselves.

Controlled unitaries

A general controlled unitary is represented in the following form in quantum circuit diagrams:



This can be read as saying that, when the first qubit equals 1, then the unitary on the second qubit is implemented. It is easy to see that this has block matrix representation

$$\begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} \tag{51}$$

where each element is a 2×2 matrix.

One of the steps to achieving our ultimate goal of showing that CNOT, H, T is a universal gate-set requires that controlled gates can be decomposed into circuits containing only CNOTs and single-qubit unitaries, so we now show this. We first note that any unitary U can be represented in the form:

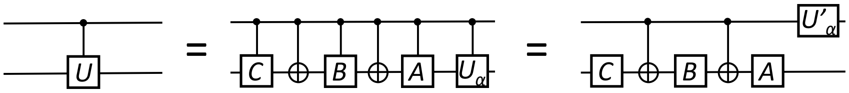
$$U = e^{i\alpha} AXBXC \tag{52}$$

where $ABC = I$. To see this, we set:

$$A = R_z(b) R_y(c/2) \tag{53}$$
$$B = R_y(-c/2) R_z(-(d+b)/2) \tag{54}$$
$$C = R_z((d-b)/2) \tag{55}$$

from which it can be verified that $ABC = I$, and $AXBXC = R_z(b)R_y(c)R_z(d)$, i.e., $U = e^{i\alpha} R_z(b) R_y(c) R_z(d)$ which therefore suffices to represent any single-qubit unitary, by (47).

Letting $U_\alpha = \begin{bmatrix} e^{i\alpha} & 0 \\ 0 & e^{i\alpha} \end{bmatrix}$ and $U'_\alpha = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{bmatrix}$, we can see that using this general unitary decomposition we have:



That is, we can decompose U according to (52) – and because we are considered controlled-U, we can simply control each of the matrices in the product individually. We can drop the control for the three matrices A, B, C because: when the top qubit is equal to one, operations on both qubits are executed, as shown; and when the top qubit is 0, then neither of the CNOTs do anything, and so the second qubit is multiplied by $ABC = I$, which does nothing – so it does not matter that we have dropped the control and A, B and C are unconditionally executed on the second qubit. We can also see that the controlled unitary CU_α can be replaced with an uncontrolled U'_α on the first qubit because:

$$CU_\alpha = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{i\alpha} & 0 \\ 0 & 0 & 0 & e^{i\alpha} \end{bmatrix} = U'_\alpha \otimes I \tag{56}$$

Thus we have shown how a general controlled unitary, U, can be executed using a circuit consisting only of CNOT gates and single-qubit unitaries.

Any n-qubit unitary can be decomposed into a product of "two-level" unitaries

In the next step towards showing that CNOT, H, T is a universal gate-set, we show that any n-qubit unitary (i.e., represented by a $N \times N$ element matrix, where $N = 2^n$) can be decomposed into a product of "two-level" unitaries. A N-element two-level unitary, U, is a unitary matrix whose elements are all equal to the corresponding element in the identity matrix of the same dimension, except for in up to four places: U_{ii}, $U_{ij}\, U_{ji}\, U_{jj}$ for some $1 \leq i, j \leq N$. First, we consider a general N element unitary matrix:

$$W = \begin{bmatrix} W_{11} & W_{12} & W_{13} & \dots \\ W_{21} & W_{22} & W_{23} & \\ W_{31} & W_{32} & W_{33} & \\ \vdots & & & \ddots \end{bmatrix} \tag{57}$$

We now give a procedure for pre-multiplying W by 2-level unitary matrices, such that the 2^{nd} to N^{th} elements of the first column of the resultant matrix are all zero. To begin with, if W_{21} is non-zero, then we pre-multiply by:

$$U_1 = \frac{1}{\sqrt{|W_{11}|^2 + |W_{21}|^2}} \begin{bmatrix} W^*_{11} & W^*_{21} & 0 & \dots \\ W_{21} & -W_{11} & 0 & \\ 0 & 0 & 1 & \\ \vdots & & & \ddots \end{bmatrix} \tag{58}$$

If W_{21} *is* zero, we skip this step – which we can think of as setting U_1 as the identity (note that the identity is a two-level unitary). We can see that:

$$W' = U_1 W = \begin{bmatrix} W'_{11} & W'_{12} & W'_{13} & \dots \\ 0 & W'_{22} & W'_{23} & \\ W'_{31} & W'_{32} & W'_{33} & \\ \vdots & & & \ddots \end{bmatrix} \tag{59}$$

Next, if the third element of the first column of W' is non-zero we multiply by (if it is zero, again skip, or equivalently set U_2 to be the identity)):

$$U_2 = \frac{1}{\sqrt{|W'_{11}|^2 + |W'_{31}|^2}} \begin{bmatrix} W^*_{11} & 0 & W^*_{31} & \dots \\ 0 & 1 & 0 & \\ W'_{31} & 0 & -W'_{11} & \\ \vdots & & & \ddots \end{bmatrix} \tag{60}$$

Which gives us:

$$W'' = U_2W' = U_2U_1W = \begin{bmatrix} W''_{11} & W''_{12} & W''_{13} & W''_{14} & \cdots \\ 0 & W''_{22} & W''_{23} & W''_{14} & \\ 0 & W''_{32} & W''_{33} & W''_{34} & \\ W''_{41} & W''_{42} & W''_{43} & W''_{44} & \\ \vdots & & & & \ddots \end{bmatrix} \tag{61}$$

Continuing in this manner, we eventually get:

$$W''' = U_{N-1} \times \cdots \times U_2U_1W = \begin{bmatrix} W'''_{11} & W'''_{12} & W'''_{13} & W'''_{14} & \cdots \\ 0 & W'''_{22} & W'''_{23} & W'''_{24} & \\ 0 & W'''_{32} & W'''_{33} & W'''_{34} & \\ 0 & W'''_{42} & W'''_{43} & W'''_{44} & \\ \vdots & & & & \ddots \end{bmatrix} \tag{62}$$

However, we know that W''' is unitary, which means that W'''_{11} has modulus 1, and the remainder of the first row of W''' must equal zero, we thus pre-multiply by:

$$U_N = \begin{bmatrix} (W'''_{11})^* & 0 & 0 & \cdots \\ 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ \vdots & & & \ddots \end{bmatrix} \tag{63}$$

This can be seen as a special case of a two-level matrix, as per the earlier definition. So we get that:

$$\tilde{W} = U_N \times \cdots \times U_2U_1W = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & \tilde{W}_{22} & \tilde{W}_{23} & \tilde{W}_{14} & \\ 0 & \tilde{W}_{32} & \tilde{W}_{33} & \tilde{W}_{34} & \\ 0 & \tilde{W}_{42} & \tilde{W}_{43} & \tilde{W}_{44} & \\ \vdots & & & & \ddots \end{bmatrix} \tag{64}$$

Which we rearrange to give:

$$W = U_1^\dagger \times \cdots \times U_N^\dagger \tilde{W} \tag{65}$$

It is easy to see that the inverse of any two-level unitary must itself be a two-level unitarity, so we have now expressed W as a product of two-level unitaries and $\tilde{W}$. However, $\tilde{W}$ is of the form of the block matrix:

$$\tilde{W} = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{W}' \end{bmatrix} \tag{66}$$

i.e., from (64), and so we can decompose $\tilde{W}'$ in the same manner as given above. Thus, we can decompose any unitary into a product of two-level unitaries.

Any two-level unitary can be expressed in terms of single-qubit unitaries and CNOTs

Following this, we now show that any two-level unitary can be implemented using single-qubit unitaries and CNOT gates. Firstly, we have that the general form of a two-level matrix is such that it operates non-trivially on (at most) two of the computational basis states. For example, the two-level three-qubit unitary:

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & U_{11} & 0 & 0 & 0 & 0 & U_{12} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & U_{21} & 0 & 0 & 0 & 0 & U_{22} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{67}$$

sends $|001\rangle \rightarrow U_{11}|001\rangle + U_{21}|110\rangle$, and $|110\rangle \rightarrow U_{12}|001\rangle + U_{22}|110\rangle$, and all other basis states are unchanged. For convenience it is helpful to define the unitary:

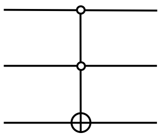
$$U' = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \tag{68}$$

The essential idea is to implement any two-level unitary with only CNOTs and single-qubit unitaries by using multi-controlled-not gates to "pick out" the computational basis states upon which U does not act trivially, and thus re-order the computational basis states such that these differ in only one bit, and then apply an appropriately controlled unitary. This is conventionally achieved by a *Gray code* which is a series of binary strings connecting two binary strings of the same length, such that upon each string-transition, only a single-bit changes. It is trivial to see that a Gray code with at most $n-1$ bits always exists to connect any two n-bit strings (as the strings to be connected can differ in a maximum of n places).

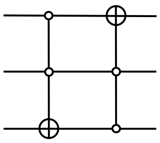
The procedure for using Gray codes to demonstrate that CNOTs and single-qubit unitaries can be used to implement any two-level unitary is better understood through an illustrative example, rather than a comprehensive, rigorous proof. Taking the above example, we first define a Gray code connecting the two states upon which U acts non-trivially (001 and 110):

001
000
100
110

We now use the Gray code to re-order the computational basis states. Firstly, by performing the multi-controlled gate:



Where (as is common notation) the white circle denotes that the operation is controlled such that it is enacted when the control bit is 0, not 1. This gate sends $|001\rangle \rightarrow |000\rangle$ (and also $|000\rangle \rightarrow |001\rangle$, although this is unimportant). Proceeding in this way, we use the remainder of the Gray code to set up control gates such that the overall effect is to send $|001\rangle \rightarrow |100\rangle$ (that is, the penultimate state in the Gray code):

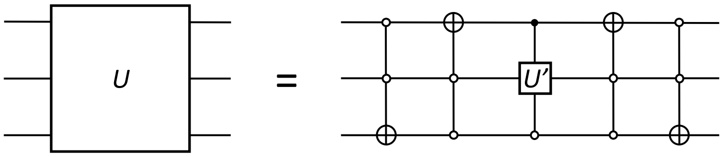


In fact we can see that the effect on all of the computational basis states is:

$$\begin{aligned}
|000\rangle &\to |001\rangle \to |001\rangle \\
|001\rangle &\to |000\rangle \to |100\rangle \\
|010\rangle &\to |010\rangle \to |010\rangle \\
|011\rangle &\to |011\rangle \to |011\rangle \\
|100\rangle &\to |100\rangle \to |000\rangle \\
|101\rangle &\to |101\rangle \to |101\rangle \\
|110\rangle &\to |110\rangle \to |110\rangle \\
|111\rangle &\to |111\rangle \to |111\rangle
\end{aligned}$$

In order for some computational basis state to be sent to another by a network of multi-controlled-nots of this form, it is necessary that, for some multi-controlled-not all but one of the bits are "picked out" by the controls and then the other bit is flipped. Notably, in the above example, the final state of the Gray code, $|100\rangle$, has not been changed. We can see that this will be true in general, as the final state will be no closer than a Hamming distance of 2 from any of the Gray code states that are "picked out" (that is, because the penultimate Gray code state is not picked out), and there is no possibility of it being altered. As the network of multi-controlled-not gates is reversible, it follows that no other computational basis state will be sent to the final Gray code state.

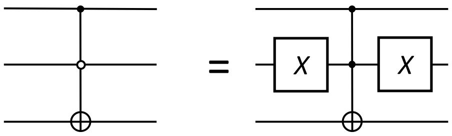
The overall action of this construction is to leave the final Gray code state unaltered, whilst sending the initial one to a computational basis state that differs in only one bit from the final state. It follows that we can use (in this example) the first and third qubits as control qubits for the unitary U' applied to the second qubit. Putting this all together, along with the same multi-controlled not circuit at the end to unscramble the computational basis vectors, we get:



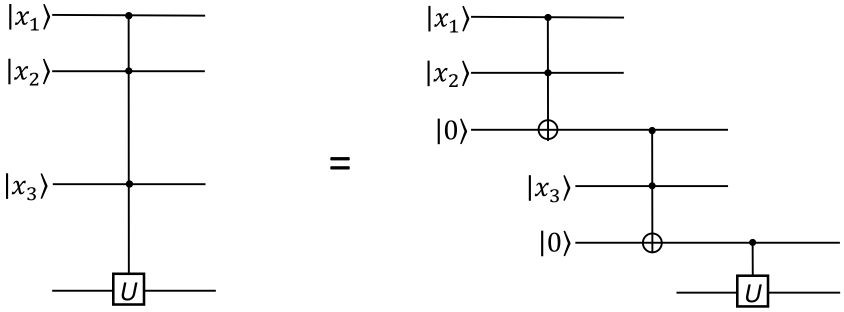
Where we also must use the previous decomposition of any controlled single-qubit unitary into CNOTs and single-qubit unitaries (i.e., applied to the controlled-U'). We have thus given a general method to implement two-level unitaries with only single-qubit unitaries and CNOT gates. Putting this together with the results above, we can see that this implies that CNOT gates and single-qubit unitaries are universal for quantum computation, subject to a few further details:

- One subtlety to be aware of is that our definition of a two-level unitary is such that it covers the case of a matrix which is the identity except for a single element on the leading diagonal (which must still have modulus one, obviously). This is the case for U_N, above, for example and can be implemented using the above construction by considering the two-level unitary to act on the computational basis state corresponding to the non-1 term on the leading diagonal, and any other computational basis state (in practise, this may not be efficient, but suffices to show that this method applies to all two-level unitaries).

- A controlled operation which is such that the operation is enacted when the control bit is equal to zero can easily be transformed into one controlled when the control bit is equal to 1, by the introduction of two X gates, e.g.



- Additionally, a multi-controlled unitary can easily be expressed in terms of Toffoli gates and a singly-controlled unitary (so long as we have access to some ancillas), eg.



And we saw in lecture how a Toffoli gate can itself be decomposed into gates from CNOT, H, T.

CNOT, H, T is a universal gate-set

We are now ready to put all of the above results together to show that CNOT, H, T is a universal gate-set. As we have seen that any n-qubit unitary can be decomposed into a product of single-qubit unitaries and CNOTs, it therefore suffices to show that H and T can generate any single-qubit unitary. The key to doing so is to use the earlier result that any single-qubit unitary can be decomposed into three rotations around two fixed (non-parallel) axes. Clearly, as rotations are operations defined by continuously varying parameters (i.e., the angle of rotation), we cannot hope to *exactly* synthesise arbitrary single-qubit unitaries from a finite gate-set, but we will instead show we can get “epsilon close”.

The full proof is rather involved, and requires a number of mathematical preliminaries which are beyond the scope of these notes. It is, however, possible to gain a solid appreciation of how H and T can be combined to give an epsilon-close approximation of any single-qubit unitary. First, consider the operator $THTH$, which is best thought of as HTH followed by T. T is a rotation of

$\pi/4$ radians around the z axis of the Bloch sphere, and HTH is a rotation of $\pi/4$ radians around the x axis of the Bloch sphere. So it follows that:

$$\begin{aligned} T(HTH) &= \exp(-i(\pi/8)Z)\exp(-i(\pi/8)X) \\ &= (\cos(\pi/8)I - i\sin(\pi/8)Z)\,(\cos(\pi/8)I - i\sin(\pi/8)X) \\ &= \cos^2(\pi/8)I - i\left[\cos(\pi/8)X + \sin(\pi/8)Y + \cos(\pi/8)Z\right]\sin(\pi/8) \end{aligned} \tag{69}$$

Crucially, this is a rotation through an angle θ such that $\cos\theta/2 = \cos^2\pi/8$ which is an irrational fraction of 2π around an axis in the direction $(\cos\pi/8, \sin\pi/8, \cos\pi/8)$. Because the angle of rotation is an irrational fraction of 2π, and rotation is an operation that is effectively modulo 2π the *pigeonhole principle* can be used to assert that a sufficiently large number of applications of $THTH$ can be used to obtain a rotation that is satisfactorily close to some target rotation angle.

We apply the same argument for the operator, $HTHT$, which is best thought of the above operator, $THTH$ pre- and post-multiplied by H, and it can be shown that this performs a rotation that is an irrational fraction of 2π around an axis in direction $(\cos\pi/8, -\sin\pi/8, \cos\pi/8)$. As the axes that $HTHT$ and $THTH$ rotate around are distinct, we can use (50) to assert that *any* unitary can be decomposed into three rotations around these two axes (up to an unimportant global phase), and moreover we have just seen that repeated applications of $HTHT$ and $THTH$ serve to achieve any desired angle (to a specified accuracy). Thus we have that any arbitrary single-qubit unitary can be decomposed as:

$$U = (THTH)^a(HTHT)^b(THTH)^c \tag{70}$$

for some integers a, b and c. This proof sketch has been vague about exactly what is meant by being "epsilon close": implicitly, in the above analysis, we have taken this to mean that we can rotate to within an angle δ of the target angle. A standard measure for closeness of two quantum operators, U, V, is the error:

$$E(U,V) = \max_{|\psi\rangle} ||(U - V)\,|\psi\rangle\,|| \tag{71}$$

We let $\hat{n}$ be a unit vector in the direction $(\cos\pi/8, \sin\pi/8, \cos\pi/8)$, and repeat $R_{\hat{n}}(\theta)$ a number, n, of times such that the difference between some target rotation angle α and $n\theta \mod 2\pi$ is such that $|\alpha - n\theta \mod 2\pi| \leq \delta$. We can always choose δ, and hence n such that

$$E((R_{\hat{n}}(\theta))^n, R_{\hat{n}}(\alpha)) \leq \frac{\epsilon}{3} \tag{72}$$

The factor of 1/3 is needed, because it can also be shown that when approximate unitaries are multiplied, the error at worst adds – and so, noting that the arbitrary unitary has been decomposed into three rotations (i.e., as in (70)), in this way we can achieve an overall error of ϵ at worst. To complete this sketch of the proof, we note two further pertinent points:

1. As the error is additive at worst, a sequence of n_u approximate unitaries, each of which is ϵ close to the target unitary, incurs an error that is $n_u\epsilon$ at worst. Such linear growth in the error is acceptable.

2. What we really care about is not the error given in (71) but rather the discrepancy in measurement outcome probabilities. However, it is possible to show that the error in measurement outcome is at most $2E(U,V)$ for an arbitrary measurement.

5.3 Compilation of quantum circuits

Knowing that some chosen gate-set is universal is not the same as knowing how to represent any unitary as a quantum circuit using only gates from that gate-set. In general, a simple counting argument can be used to show that not *all* n-qubit unitaries can be efficiently (in poly(n) gates) executed in a quantum circuit with gates chosen from some specified universal gate-set. However, as noted in the lectures, the Solovay-Kitaev theorem (the proof of which is beyond the scope even of these supplementary notes) does assert that *any* universal gate-set is sufficient to efficiently execute a unitary that has already been decomposed into CNOTs and single-qubit gates. However, in real-world situations the Solovay-Kitaev theorem, although constructively proven, does not provide a practical way to synthesise a quantum circuit from a certain universal gate-set.

Quantum compilation, broadly speaking, concerns the process of building an efficient quantum circuit that can be executed on a certain target hardware (i.e., with a certain gate-set, qubit connectivity and other properties). This entails, amongst other things, synthesising a circuit that achieves the desired function but uses only gates from the specified gate-set; performing optimisations on the circuit (i.e., combining gates to find a functionally equivalent, but simpler circuit – similar in spirit to Boolean algebra simplification in classical logic); and *qubit routing* – the process of inserting swaps into the circuit to enable it to run on the target hardware.

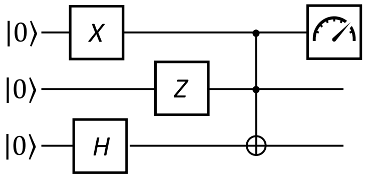
Unlike in its classical counterpart, in quantum compilation we are aided by the fact that the input to the compiler is itself typically a quantum circuit (rather than a higher-level language – although various efforts are afoot to do this). This is because for most quantum algorithms, even those which will potentially yield spectacular results, it is reasonable to write down the quantum circuit explicitly. This means that the compilation may amount to re-writing the circuit into an efficiently executable form, which is itself a highly non-trivial task that requires a great deal of understanding of quantum computing but is to some extent helped by the fact that many quantum algorithms are already described in terms of primitive quantum gates, namely CNOT, H, T, X, Y, Z, S – the latter four of which we saw in the lecture how to express in terms of CNOT, H, T. Of course, there are exceptions, such as the aforementioned rotation gates, which are commonly used in *variational* or *parameterised* quantum circuits.

5.4 Implicit and deferred measurement

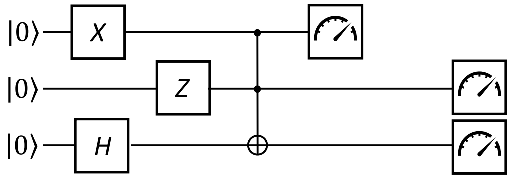
In the lecture we defined a quantum circuit as some unitary operation (represented as some quantum gates) acting on the state $|0\rangle$ from which some qubits are then measured. This definition raises two possible questions: What about mid-circuit measurement? And what about the unmeasured qubits? Fortunately, there are principles that deal with each of these, and thus provide techniques that are sometimes helpful in the analysis of quantum circuits.

Principle of implicit measurement. *Any unmeasured qubits can be treated as being measured.*

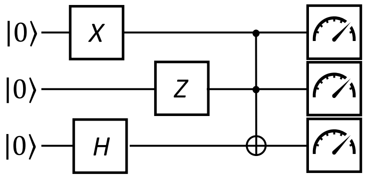
For example, if this is our quantum circuit:



Then by causality it is irrelevant whether or not we *later* decide to measure the remaining qubits:



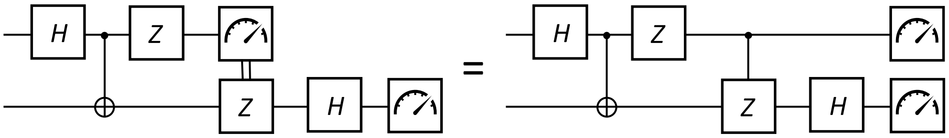
But in the quantum circuit model, this is equivalent to the later measurements happening at the same time:



Formally, this can be proven by showing that *reduced density matrices* are insensitive to whether or not the other qubits are measured. The principle of implicit measurement is very similar to the no-signalling principle: if we measure some qubits of an entangled state, then the statistics of those measurements are unchanged by whether or not the other qubits are measured.

Principle of deferred measurement. *Any mid-circuit measurement can be deferred to the end of the circuit, and classical-control conditioned on the measurement outcome replaced by quantum control.*

That is, for example:



The principle of deferred measurement can be reasoned along the lines: *whatever happens in the rest of the circuit must be consistent with the measurement outcome observed*, and thus it is irrelevant whether that control is classical or quantum with the measurement outcome deferred.

It is worth noting that, although the principle of deferred measurement is valid as a tool for circuit analysis for *any* quantum circuit, its use may mean that the physical interpretation is lost. For example, consider quantum teleportation: using the principle of deferred measurement one can obtain the correct final quantum state, but one loses the crucial factor that Alice only sends classical information to Bob.

5.5 Post selection

One final piece of quantum circuit jargon that its useful to be aware of is "post-selection". Post selection simply means that if we run a quantum circuit, we assume that we get whatever measurement outcome we choose. So, for example, if we have a quantum circuit where we measure a single qubit in the computational basis, with probability of measurement outcome $|0\rangle$ and $|1\rangle$ each equal to one-half, then we may say that we "post select" on the outcome $|0\rangle$. The implication is that when physically running the circuit, we may to repeat a number of times until we get the outcome we want.

In the case given above, clearly we can expect that we will not have to repeat many times until we get the outcome $|0\rangle$, however it is worth being aware that if we post-select some n-qubit outcome, then if we were to actually run our circuit the expected number of repeats to obtain the post-selected outcome grows exponentially in n.

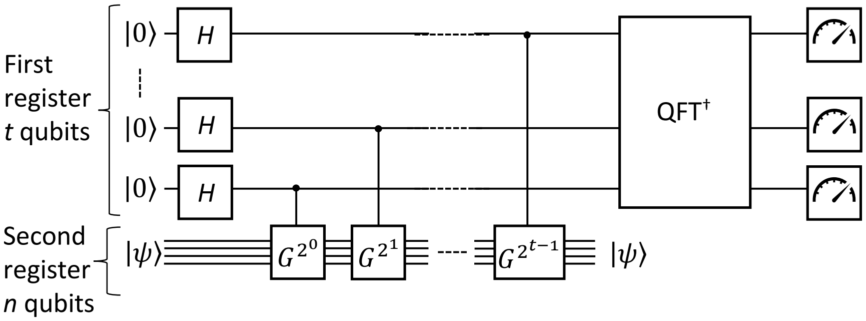
At first glance, post-selection seems a slightly odd idea – however when analysing circuits that include measurements (especially mid-circuit measurements), it can often be convenient to post-select the desired outcome, and then account for the (expected) number of repeats to obtain that outcome later on. Moreover, using post-selected complexity classes can sometimes provide a surprisingly powerful analytical tool in computational complexity theory – even when deriving results for non post-selected classes.

9 Quantum Fourier Transform & Quantum Phase Estimation

9.1 Quantum counting

In Lecture 8 we saw how Grover's algorithm enables us to find a single marked entry in an unstructured search problem, and argued that this method extents trivially to some *known* number, M, of marked solutions. But what about when M is unknown, as would typically be the case if Grover search were to be used to solve an NP-complete problem? It turns out we can combine Grover search with the quantum Fourier transform to (approximately) *count* the number of solutions in an unstructured search problem, thus finding M.

If we let the Grover iterate be G, i.e., $G = (W \otimes I)V$, then this is achieved by applying the following circuit, from which M can be found from the resultant eigenvalue estimate.



To see how this works, first note from the lecture that the effect of the Grover iterate is to rotate by an angle 2θ in the plane spanned by $|y\rangle$ (an equal superposition of unmarked states), and $|z\rangle$ (an equal superposition of marked states). That is, if we consider the projection onto this plane, and define $|y\rangle = [1, 0]^T$ and $|z\rangle = [0, 1]^T$, we have that:

$$G = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix} \tag{73}$$

from which we can verify that G has an eigenvalue $e^{2i\theta}$ with eigenvector $(1/\sqrt{2})[i, 1]^T = (1/\sqrt{2})(i\,|y\rangle + |z\rangle)$ and also $e^{i(2\pi - 2\theta)}$ with eigenvector $(1/\sqrt{2})[-i, 1]^T = (1/\sqrt{2})(-i\,|y\rangle + |z\rangle)$. Letting these eigenvectors be $|e_1\rangle$ and $|e_2\rangle$, respectively, we can see that the uniform superposition, $|\psi\rangle$, can be expressed as a superposition of $|e_1\rangle$ and $|e_2\rangle$ (using definitions from Lecture 8, adjusted for the case in which we have M marked entries):

$$\begin{aligned}
|\psi\rangle &= \frac{1}{\sqrt{N}} \sum_{x\in\{0,1\}^n} |x\rangle \\
&= \frac{\sqrt{N-M}}{\sqrt{N}} \left(\frac{1}{\sqrt{N-M}} \sum_{x \text{ s.t.} f(x)=0} |x\rangle \right) + \frac{\sqrt{M}}{\sqrt{N}} \left(\frac{1}{\sqrt{M}} \sum_{x \text{ s.t.} f(x)=1} |x\rangle \right) \\
&= \frac{\sqrt{N-M}}{\sqrt{N}} |y\rangle + \frac{\sqrt{M}}{\sqrt{N}} |z\rangle \\
&= \frac{\sqrt{N-M}}{\sqrt{N}} \frac{i}{\sqrt{2}} (|e_2\rangle - |e_1\rangle) + \frac{\sqrt{M}}{\sqrt{N}} \frac{1}{\sqrt{2}} (|e_1\rangle + |e_2\rangle) \qquad (74)
\end{aligned}$$

Therefore the QPE will return an estimate of either 2θ or $2\pi - 2\theta$ (i.e., the eigenvalues of $|e_1\rangle$ and $|e_2\rangle$) if we prepare the second register in the state, $|\psi\rangle = 1/\sqrt{N} \sum_{x\in\{0,1\}^n} |x\rangle$, that is the equal superposition of all computational basis states, as prepared by a single layer of Hadamard gates.

From either of these possibilities, it is possible to extract an estimate of θ/π (the factor of π owing to the definition of phase) with maximum error 2^{-t} from the QPE procedure (note that strictly speaking some additional qubits are required to mitigate the possibility of QPE failing, however as the number of qubits required is merely logarithmic in the reciprocal of the tolerable failure probability, we omit this from simplicity).

We can now analyse how much this round-off error in the phase estimation affects our estimate of M. Specifically, we will express the error in the estimate of M, ϵ_M, in terms of the error in the estimate of θ, ϵ_θ, and hence t. Firstly, following the analysis in Lecture 8, if there are M (rather than one) marked states, we have that $\sin\theta = \sqrt{M/N}$. Thus we have:

$$\frac{\epsilon_M}{N} = \sin^2(\theta + \epsilon_\theta) - \sin^2(\theta) \qquad (75)$$

We can re-arrange this as:

$$\frac{\epsilon_M}{N} = (\sin(\theta + \epsilon_\theta) - \sin(\theta)) (\sin(\theta + \epsilon_\theta) + \sin(\theta)) \qquad (76)$$

which allows us to bound the maximum absolute value of ϵ_M:

$$\frac{|\epsilon_M|}{N} = |\sin(\theta + \epsilon_\theta) + \sin(\theta)| \times |\sin(\theta + \epsilon_\theta) - \sin(\theta)| \qquad (77)$$

As the magnitude of the gradient of $\sin\theta$ is at most one, the second term in the right-hand side can be upper-bounded by $|\epsilon_\theta|$. Using a simple result from trigonometry $|\sin(\theta+\epsilon_\theta)| < \sin\theta + |\epsilon_\theta|$ (noting that the problem set-up is such that $0 \leq \theta \leq \pi/2$, so $\sin\theta$ is non-negative) and also $M/N = \sin^2\theta$ and $\epsilon_\theta \leq \pi 2^{-t}$ (the factor of π because we are performing phase estimation, so estimate θ/π) we get:

$$\begin{aligned}
\frac{|\epsilon_M|}{N} &< |2\sin\theta + \epsilon_\theta| \times |\epsilon_\theta| \\
&< \left(2\sqrt{M/N} + \frac{\pi}{2^t} \right) \frac{\pi}{2^t} \\
\implies |\epsilon_M| &< \left(2\sqrt{MN} + \frac{\pi N}{2^t} \right) \frac{\pi}{2^t} \qquad (78)
\end{aligned}$$

finally, choosing $t = n/2$, i.e., $2^t = \sqrt{2^n} = \sqrt{N}$ (for simplicity we let n be even) we get:

$$\begin{aligned} |\epsilon_M| &< \left(2\sqrt{MN} + \pi\sqrt{N}\right)\frac{\pi}{\sqrt{N}} \\ &= \pi\left(2\sqrt{M} + \pi\right) \end{aligned} \tag{79}$$

So we have that we can estimate M with maximum error $\mathcal{O}(\sqrt{M})$, and we can see that for $t = n/2$ we must call the Grover operate G a total of $2^{n/2} = \sqrt{N}$ times. Therefore we only require $\mathcal{O}(\sqrt{N})$ Grover iterations to approximately count M. As previously noted, classically we would require N operations to do this exactly, and it turns out that we would still need $\mathcal{O}(N)$ operations to approximately classically count M to the same accuracy as we have achieved quantumly.

Quantum counting is closely related to quantum Monte-Carlo estimation, which also exhibits a quadratic speed-up over its classical counterpart. To see this is even easier, as we directly have that, for precision 2^{-t} we require 2^t uses of G – therefore to achieve some desired error, ϵ, we require $\mathcal{O}(1/\epsilon)$ operations, whereas classically we would require $\mathcal{O}((1/\epsilon)^2)$. Perhaps surprisingly, we can see that this quadratic quantum speed-up is in terms of the desired accuracy (reciprocal of the tolerable error), and so applies to problems of any size.

12 Quantum Complexity

12.1 Classical simulation of quantum circuits and the Gottesman-Knill theorem

Quantum computers exploit the properties of quantum mechanics, and hence are physically quite different from classical computers. However, we have seen that quantum computing does not violate the Church-Turing thesis, and thus in *principle* any quantum algorithm is also classically computable. In fact, it is possible to directly *simulate* the behaviour of a quantum computer by performing linear algebra. In general, if we have a n-qubit algorithm this will entail multiplication of matrices of size $N = 2^n$, and it is this exponentiation that means that we consider the classical simulation of quantum algorithms to be infeasible in *practise*. There are, however, various results showing that certain classes of quantum algorithm can be efficiently simulated classically, and the most famous of these is the Gottesman-Knill theorem. The Gottesman-Knill theorem states that any quantum circuit consisting only of *Clifford group* operations (and measurements) can be classically simulated with only Poly(n) space and time resources. The Clifford group is generated by the H, S and CNOT operations – this means that the Gottesman-Knill theorem applies to quantum circuits consisting of these three gates (and other gates that can be made out of combinations of these three, such as $Z = S^2$). The Gottesman-Knill theorem was formulated in terms of the *Stabiliser formalism*, and this remains how it is normally proved. The Stabiliser formalism is an alternative way of expressing quantum states, and is beyond the scope even of these supplementary notes, however the Gottesman-Knill theorem can also be proved by representing quantum states in terms of superpositions of computational basis vectors, as we conventionally do.

This proof of the Gottesman-Knill theorem is rather long-winded and technical, however it is possible to get the gist of it by noting that the Clifford group operations are such that in a Clifford circuit a single qubit can either be in the state $|0\rangle$ or $|1\rangle$, or an equal superposition of $|0\rangle$ and $|1\rangle$. Moreover, in the latter case, the phase of the $|1\rangle$ state relative to the $|0\rangle$ state in the superposition can only be one of $\{1, -1, i, -i\}$. To see this, consider the actions of any combination of H and S on a single qubit initially in the state $|0\rangle$. We also have the CNOT, which is an entangling gate, however it turns out that we can always express any Clifford circuit state in the form:

$$|\psi\rangle = \frac{1}{\sqrt{2^r}} \sum_{\mathbf{x}\in\{0,1\}^n} i^{(\mathbf{p}^T\mathbf{x}+s)}(-1)^{\mathbf{x}^T Q\mathbf{x}} |A\mathbf{x}+\mathbf{t}\rangle \tag{80}$$

where $\mathbf{x}$ is taken to be a column vector; $\mathbf{t} \in \mathbb{Z}_2^n$; $\mathbf{p} \in \mathbb{Z}_2^n$; $Q \in \mathbb{Z}_2^{n\times n}$; $A \in \mathbb{Z}_2^{n\times n}$; $s \in \mathbb{Z}_2$, and addition is modulo-2. That is, all of the terms are binary vectors and matrices of size $n \times 1$ or $n \times n$. This can be thought of in the same way that we thought of single-qubit Clifford states (i.e., they are $|0\rangle$ or $|1\rangle$, or an equal superposition of $|0\rangle$ and $|1\rangle$ where the phase of the $|1\rangle$ state relative to the $|0\rangle$ state is one of $\{1, -1, i, -i\}$), but now the superpositions can be entangled superpositions (i.e., over multiple qubits).

Thus, if we have a Clifford circuit, we can represent the state at any point part way through the circuit in the form of (80) and this requires an amount of memory which is only polynomial in the number of qubits n (i.e., to store $\mathbf{t}$; $\mathbf{p}$; Q; A; and s). It can also be shown that to update the state represented in the form (80) when any Clifford group operation is applied to any qubit requires a number (classical) operations which is also only polynomial in n. So it follows that any Clifford circuit can be simulated classically, by working through the circuit on a classical computer, updating $\mathbf{t}$; $\mathbf{p}$; Q; A; and s after each operation, and thus efficiently keeping track of the state. This requires only time and space that is polynomial in n, and contrasts with general quantum states which correspond to matrices of size $N = 2^n$, and thus naive simulation of which (simply by applying matrix multiplications to simulate each gates) would require time and space that is exponential in n.

There are various more efficient ways to classically simulate general quantum circuits, however it remains the case that it is believed that this will take an exponential (in n) amount of time for at least some circuits. Thus the Gottesman-Knill theorem is very significant, as it tells us that quantum circuits must contain at least some non-Clifford operations to be non-classically simulable (in polynomial time). This result was originally somewhat surprising, as Clifford circuits can generate very highly-entangled quantum states. Thus we can conclude that, whilst necessary for quantum advantage, entanglement is not sufficient – and that the computational power of quantum mechanics is quite a subtle thing.

16 Case Studies in Near-term Quantum Computation

16.1 HHL (sketch)

As we have seen throughout the course, quantum circuits with n qubits essentially amount to operations on matrices of size $N = 2^n$. With this in mind, it may appear that the most obvious application of quantum computing is to perform linear algebra on systems of size N' using only $n' = \lceil \log N' \rceil$ qubits, so achieving an exponential speed-up directly. So it is perhaps a little surprising that the discovery of a quantum algorithm to solve a system of linear equations came relatively late on in the formative years of quantum computing: in 2009 by Aram Harrow, Avinatan Hassidim, and Seth Lloyd – whose initials give the algorithm its common name "HHL".

HHL addresses the following problem: let $\mathbf{x}$ be a N element vector, $\mathbf{b}$ be a M element vector and A be a $M \times N$ matrix such that:

$$\mathbf{b} = A\mathbf{x} \tag{81}$$

where A and $\mathbf{b}$ are known, and we want to find $\mathbf{x}$. Classically, we must use matrix inversion (finding the pseudo-inverse if A is not square), which requires $\mathcal{O}(N^3)$ operations (assuming $M \in \Theta(N)$). If A is sparse, i.e., each row of A has only $s << N$ non-zero entries, this can be reduced to $\mathcal{O}(N \log N)$

operations.

Quantumly, we can do much better. First, we tweak the equation such that the matrix is Hermitian. We can see that we can find $\mathbf{x}$ by solving:

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} 0 & A \\ A^\dagger & 0 \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{x} \end{bmatrix} \tag{82}$$

which we can see is an equation of the form:

$$\tilde{\mathbf{b}} = \tilde{A}\tilde{\mathbf{x}} \tag{83}$$

where $\tilde{A}$ is Hermitian. This corresponds to a system of $\tilde{N} = N + M$ equations, and we let $\tilde{N} = 2^{\tilde{n}}$ be a power of 2 for simplicity. To execute HHL, we first must encode $\tilde{\mathbf{b}}$ as a $\tilde{n}$ qubit quantum state, $|b\rangle$. This we do using *amplitude encoding*, that is such that $|b\rangle$ is superposition of $2^{\tilde{n}}$ computational basis states, with the co-efficient of each basis state equal to the corresponding element of $\tilde{\mathbf{b}}$:

$$|b\rangle = \frac{1}{\sum_i |\tilde{\mathbf{b}}_i|^2} \sum_{i=1}^{\tilde{N}} \tilde{\mathbf{b}}_i \, |i\rangle \tag{84}$$

where $|i\rangle$ is the i^{th} computational basis state. It should be noted that this process of "loading" $\tilde{\mathbf{b}}$ into the quantum computer will not always be efficiently possible. However Harrow, Hassidim, and Lloyd argue that it will be for some interesting problems, and moreover HHL also applies to solving linear systems of equations where the data is already in quantum form (i.e., because the data is from some quantum sensing process, or if HHL is called as a subroutine by some other quantum algorithm).

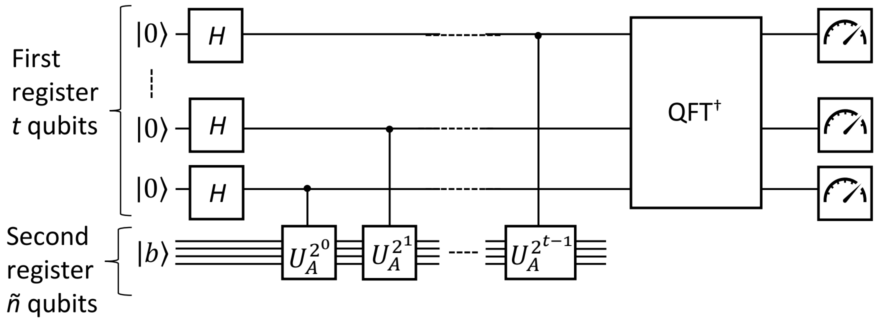
If $\tilde{A}$ is sparse, then it is possible to efficiently construct a quantum circuit to apply the unitary:

$$U_A = e^{2\pi i \tilde{A}} \tag{85}$$

and we also note that $|b\rangle$ can be expressed as a superposition of the eigenvalues of A,

$$|b\rangle = \sum_{i=1}^{\tilde{N}} \beta_i \, |e_i\rangle \tag{86}$$

where β_i are co-efficients, and $|e_i\rangle$ are eigenvectors of $\tilde{A}$. We therefore have all of the ingredients necessary to apply quantum phase estimation to $|b\rangle$ in the basis of $\tilde{A}$.



(note that t can be chosen such a that both the overall computational complexity and error are satisfied). This yields the state:

$$|\psi\rangle = \sum_{i=1}^{\tilde{N}} \beta_i \, |\lambda_i\rangle \, |e_i\rangle \tag{87}$$

where λ_i is the eigenvalue of $\tilde{A}$ corresponding to the i^{th} eigenvector (i.e., because the *phases* of the eigenvalues of U_A are the eigenvalues of $\tilde{A}$). The next step of HHL is to apply the non-unitary operation:

$$\sum_{i=1}^{\tilde{N}} \beta_i \, |\lambda_i\rangle \, |e_i\rangle \to \sum_{i=1}^{\tilde{N}} \frac{C}{\lambda_i} \beta_i \, |\lambda_i\rangle \, |e_i\rangle \tag{88}$$

for some constant C. As this is non-unitary, it requires some measurements, and this step fails with some probability. However, using *amplitude amplification* (a generalisation of Grover's algorithm), this failure probability can be suppressed to be acceptably small. If we then uncompute the register containing $|\lambda_i\rangle$ we get:

$$\sum_{i=1}^{\tilde{N}} \frac{C}{\lambda_i} \beta_i \, |e_i\rangle = C\tilde{A}^{-1} \, |b\rangle = C \, |x\rangle \tag{89}$$

To see this, recall that we have a decomposed $|b\rangle$ into a superposition of eigenvectors of $\tilde{A}$, and then divided each term in the superposition by the corresponding eigenvalue of $\tilde{A}$ – which clearly amounts to multiplying $|b\rangle$ by the inverse of $\tilde{A}$.

As C is a known constant, we have therefore found $|x\rangle$, which is a quantum encoding of $\tilde{\mathbf{x}}$, again using amplitude encoding. So it follows that we cannot extract the entirety of $\tilde{\mathbf{x}}$, as that would require us to repeat HHL many times and measure the outcome to build up an estimate of $\tilde{\mathbf{x}}$ with sufficiently small uncertainty – a process known as *quantum state tomography*, which would require so many repeats that the quantum advantage would be nullified. We can, however, quickly extract "global" properties of $|x\rangle$ such as its average (amongst others) which are still hard to find classically.

HHL is an algorithm on $\Theta(\tilde{n})$ qubits (where $\tilde{n} = \log \tilde{N}$), and so we may expect that the computational complexity is some polynomial of $\tilde{n}$, and indeed this turns out to be the case. Specifically, HHL has computational complexity:

$$\tilde{\mathcal{O}}(\kappa^2 s^2 \log N/\epsilon) \tag{90}$$

where s is the sparsity (as already briefly mentioned), ϵ is the allowable error, and κ is the condition number: the ratio of the absolute value of the eigenvalue of $\tilde{A}$ with maximum absolute value, to the absolute value of the eigenvalue of $\tilde{A}$ with minimum absolute value (note that $\tilde{\mathcal{O}}$ simply means that there may be further complexity that is at worst poly-logarithmic in the terms inside the bracket, so is suppressed for simplicity). Note that $\tilde{N} \in \mathcal{O}(N)$ and so we state the asymptotic complexity in terms of the dimensions of the original matrix, A. This therefore constitutes an exponential speed-up over the classical case, even if sparsity techniques are used to reduce the computational complexity (i.e., to $\mathcal{O}(N \log N)$, as above).

So we can see that HHL indeed allows us to solve systems of linear equations exponentially faster using a quantum computer in some circumstances.