CS395T NOTES: QUANTUM COMPLEXITY THEORY

ARUN DEBRAY
SEPTEMBER 12, 2016

These notes were taken in UT Austin's CS395T (Quantum Complexity Theory) class in Fall 2016, taught by Scott Aaronson. I live-TEXed them using `vim`, so there may be typos; please send questions, comments, complaints, and corrections to a.debray@math.utexas.edu.

Contents

Lecture 1.

Introduction to Quantum Mechanics: 8/29/16

"The big secret of quantum mechanics is how simple it is once you take the physics out of it."

The course website is http://www.scottaaronson.com/qct2016/, and the syllabus is at http://www.scottaaronson.com/qct2016/syllabus-qct2016.pdf. We'll be mostly following lecture notes found at http://www.scottaaronson.com/barbados-2016.pdf.

This lecture's goal is to acquaint the listener with the basic concepts and notation that we'll use in the rest of the course; it's not presented as review, but everything else in the course depends on it. For this material, there are many excellent references, some of which are listed in the syllabus.

Quantum mechanics has a very underserved reputation for being very complicated. Mysterious, yes; counterintuitive, yes; but complicated is a bit much. All sorts of interesting consequences follow from a single change to the laws of probability, crucial to physics at the subatomic level, but thought to apply to everything in the universe.

A probability of something happening is a real number $p \in [0,1]$: it makes no sense to ask what a probability of $-1/3$ is, much less $i/3$. But quantum mechanics assigns a more general number, an *amplitude* $\alpha \in \mathbb{C}$, to an event. The thesis of quantum mechanics is that any isolated physical system's state can be described by a vector of its amplitudes.

In particular, systems in quantum mechanics have a dimension; intuitively, if there are N different things you can observe, the system is N-dimensional. The simplest quantum systems are two-dimensional, where there are two possinilities) and 1. These systems have a special name: *qubits*.

In general, we think of the state of a quantum system as a unit vector $\psi \in \mathbb{C}^N$ of length 1. These vectors are denoted using a notation that Paul Dirac invented in the 1930s, the *Dirac ket notation*. The syntax looks a little jarring at first, but is convenient in a lot of ways. A *ket* is a vector $|v\rangle$: a qubit has two basis vectors $|0\rangle$, representing an outcome of 0 and $|1\rangle$, similarly an outcome of 1, so a general state is $|v\rangle = \alpha|0\rangle + \beta|1\rangle$, representing a linear combination, or *superposition*, of the two options: $\alpha, \beta \in \mathbb{C}$ are complex numbers, and must satisfy a *normalization rule*: $|\alpha|^2 + |\beta|^2 = 1$. In other words, $|v\rangle$ stands in for the column vector $\binom{\alpha}{\beta}$. Usually, ket notation will only be for unit vectors, but sometimes we might use it more generally.

This feels a little schizophrenic. Is it both at the same time? Is it neither? In popular books, these are the only ontological categories the writer can imagine, but these really belong in a different conceptual framework altogether.

In addition to column vectors, we like row vectors too, denoted with a *bra* $\langle v|$. However, since we're in the land of complex vector spaces, taking the transpose comes along with complex conjugation, so

$\langle v| = (\alpha^* \ \beta^*)$. Combining these two notations, $\langle \cdot | \cdot \rangle$ is the notation for the inner product. Thus, that v is a unit vector is succinctly expressed in the condition $\langle v|v\rangle = 1$.

Explicitly, if $|\psi\rangle = \alpha_1|1\rangle + \cdots + \alpha_N|N\rangle$ and $|\varphi\rangle = \beta_1|1\rangle + \cdots + \beta_N|N\rangle$, their inner product is $\langle\psi|\varphi\rangle = \alpha_1^*\beta_1 + \cdots + \alpha_N^*\beta_N$. This measures how similar two vectors are: if the inner product is 1, they lie on the same line, and are related, but if it's 0, they're orthogonal, and thus very different.

Relatedly, there is an *outer product* $|\psi\rangle\langle\varphi|$, which is a rank-1 $N \times N$ matrix whose ij^{th} term is $\alpha_i\beta_j^*$.

There are two things one can do to quantum systems.

(1) One option is a *unitary transformation*. These should be thought of as doing something smooth and well-behaved. They are continuous, reversible, and deterministic.
(2) The other choice is a *measurement*. These are useful, especially if you want to actually learn anything about system, but these are discontinuous and irreversible, and famously are probabilistic. Quantum mechanics tells you probabilities, not certainties.

Maybe you're wondering how two so very different systems can coexist in the same universe. This is the *measurement problem*, and people have been discussing it for a century. In some sense, unitary transformations arise from changes of basis, but if you follow that viewpoint far enough, it seems like all of quantum mechanics is a particular change of basis! Yet there are ways in which the choice of basis matters; unitary transformations are information-preserving, relating to the very general physical principle that information cannot be destroyed. A unitary transformation might horribly transform information, but it's still there.

The measurement problem and its metaphysics notwithstanding, we can at least write down the mathematical rules for these transformations. A unitary evolution is multiplication by a matrix: $|\psi\rangle \mapsto U|\psi\rangle$, but U must be norm-preserving, so that all valid quantum states map to valid quantum states.[1] Since unitary transformations should be reversible, we'd like U to be an invertible matrix.

Exercise 1.1. Show that the following are equivalent for a linear transformation $U : \mathbb{C}^n \to \mathbb{C}^n$:

(1) U is norm-preserving and invertible.
(2) U preserves inner products, i.e. $\langle U\psi|U\varphi\rangle = \langle\psi|\varphi\rangle$ for all $|\psi\rangle, |\varphi\rangle \in \mathbb{C}^N$.
(3) $U^\dagger U = I$ (here, † denotes conjugate transpose).
(4) The rows of U are an orthonormal basis for $\mathbb{C}^N$.
(5) The columns of U are an orthonormal basis for $\mathbb{C}^N$.

Such a matrix is called a *unitary matrix*.

Example 1.2 (Qubit)**.** The simplest example is a qubit, whose vector space is spanned by two basis vectors $|0\rangle$ and $|1\rangle$ (so it has two complex dimensions, or four real dimensions). Thus, the possible superpositions are $\alpha|0\rangle + \beta|1\rangle$ such that $|\alpha|^2 + |\beta|^2 = 1$. Often, but not always, α and β will be real, making them easier to draw; in this case, we just need $\alpha^2 + \beta^2 = 1$, defining a circle.

The *plus state* is $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, and the *minus state* is $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. $(|+\rangle, |-\rangle)$ is also an orthonormal basis for this space.

What are some unitary transformations? We have the identity

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

as well as the *NOT gate*

$$\text{NOT} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

In general, a *gate* will refer to a unitary matrix that's applied to only one or a few qubits.

The identity and the NOT gate make sense for classical probability too, sending probability vectors to probability vectors. This is not true for the next matrix, called the *phase gate*:

$$\text{Phase} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

[1] This is what keeps the probabilities adding up to 1.

which sends $(\alpha, \beta) \mapsto (\alpha, -\beta)$. There are other phases, e.g. replacing -1 by another root of unity. Similarly, the *Hadamard matrix* is

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Notice that $H|0\rangle = |+\rangle$ and $H|1\rangle = |-\rangle$, but $H|+\rangle = |0\rangle$ and $H|-\rangle = |1\rangle$, so the Hadamard matrix switches the normal basis and the plus-minus basis. Thus, $H^2 = I$.

Finally, there are rotation matrices

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

This rotates counterclockwise by the angle θ.

Every unitary transformation in two dimensions is a product of rotations and reflections. Notice that the Hadamard matrix is not a rotation: we can apply $R = R_{\pi/4}$, which sends $|0\rangle \mapsto |+\rangle$ and $|1\rangle \mapsto -|-\rangle$. In general, $|\psi\rangle$ and $-|\psi\rangle$, as well as $i|\psi\rangle$, produce the same physical behavior: there's no experiment that can tell them apart. The classical analogue would be to move the whole universe twenty feet to the left: does anything actually change?

We can calculate $R|0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $R|1\rangle = (-|0\rangle + |1\rangle)/\sqrt{2}$. Evaluating on $|+\rangle$, we have to cancel out a $|0\rangle/2$ and a $-|0\rangle/2$:

$$R\frac{|0\rangle + |1\rangle}{\sqrt{2}} = \frac{\frac{|0\rangle+|1\rangle}{\sqrt{2}} + \frac{-|0\rangle+|1\rangle}{\sqrt{2}}}{\sqrt{2}} = |1\rangle.$$

This is called *destructive interference*: the two ways to obtain $|0\rangle$ were in opposite amplitudes, so they descrutively interfered, and cancelled out to zero probability. Similarly, the outcomes leading to 1 displayed *constructive interference*. A lot of the weirdness of quantum mechanics comes out of interference phenomena, since they behave so differently from classical mechanics. The *double-slit experiment* is an example: if a photon passes through two slits in an opaque material, there are alternating zones of light and dark, which makes classical sense. But the part that makes no sense classically is that if you close off one of the slits, photons can appear where they hadn't before.

This wasn't just a violation of intuition; it was a violation of the axioms of probability: classically, one assumes that the probabilities p_A and p_B of getting to that point after passing through the two slits A and B, respectively, should add to the total probability of a photon landing at that spot, but the experiment disproved that. This and its analogues in atomic nuclei, etc. are why quantum mechanics works with amplitudes instead of probabilities. In fact, the amplitude of this process is the sum of the amplitude occuring from slit A and the amplitude occuring from slit B. Destructive interference explains why closing slit B affects the answer.

Measurements. Given a qubit $\alpha|0\rangle + \beta|1\rangle$, we want to know whether it's 0 or 1. The rule is $\Pr[0] = |\alpha|^2$ and $\Pr[1] = |\beta|^2$. This is called *Born's rule*, after Max Born (who won his Nobel for work including this!). But the second, and very important, thing that happens is that the state "collapses" to whichever measurement you observed. This is much like some people one encounters: they're not certain about their opinion on a topic, but once they're asked about it, they pick an opinion and stick to it, at least until a unitary transformation is applied to them. This is why one says that measurement in quantum mechanics is an irreversible process.

In general, if we have a superposition of N outcomes $\alpha_1|1\rangle + \cdots + \alpha_N|N\rangle$, then $\Pr[i] = |\alpha_i|^2$. This is why global phase is irrelevant: the only way you can learn anything about a quantum system is measurement. Many of the paradoxes or misunderstandings people make implicitly assume there's some other way to measure the system. This also shows why there's no way to tell apart $|\psi\rangle$ and $-|\psi\rangle$: no measurement can distinguish them. It also explains interference: two amplitudes may both be nonzero, but if they're opposite in sign, the norm-squared of their sum is zero or nearly zero.

Measurement is denoted with a sort of speedometer $\Uparrow$.[2] Precomposing with a unitary transformation allows one to measure with respect to a different basis, e.g. using the Hadamard matrix is measurement in the $\{|+\rangle, |-\rangle\}$-basis. This means we'll get the outcome $+$ with probability $|\langle\psi|+\rangle|^2$ and outcome $-$ with probability $|\langle\psi|-\rangle|^2$. This is really just a rotation.

[2] This should be a semicircle with an arrow pointing to the upper right, but I don't know how to TEX that yet.

A pure 0 state always evaluates to 0. A pure 1 state always evaluates to 1. An equal superposition gives 0 half the time, and 1 half the time. But evaluating with respect to the $\{|+\rangle, |-\rangle\}$ basis turns pure states into equal superpositions and vice versa. In other words, $\Pr[|v_i\rangle] = |\langle\psi|v_i\rangle|^2$.

Example 1.3. We can generalize to systems of multiple qubits, placing them beside each other. A qubit might correspond to an electron with two energy states, or two spin directions (up and down), or any physical system that can be in either of two discrete states: quantum mechanics says there can also be a superposition. The variety of these systems leads to the variety of proposals for the physical architectures of a quantum computer.

Suppose now we have two photons. We refer to the composite of these systems with a tensor product: $(\alpha|0\rangle + \beta|1\rangle) \otimes (\gamma|0\rangle + \delta|1\rangle)$. We can distribute out the $\otimes$: the two-qubit space is actually spanned by the four basis vectors $\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\}$.[3] The amplitudes are

$$(\alpha|0\rangle + \beta|1\rangle) \otimes (\gamma|0\rangle + \delta|1\rangle) = \alpha\gamma|00\rangle + \alpha\delta|01\rangle + \beta\gamma|10\rangle + \beta\delta|11\rangle.$$

This is a unit vector in $\mathbb{C}^4$.

Conversely, one might want to factor a state as a tensor product:

$$\frac{|00\rangle - |01\rangle - |10\rangle + |11\rangle}{2} = \frac{|0\rangle - |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

However, not every state can be factored, e.g. $(|00\rangle + |11\rangle)/\sqrt{2}$: if we expanded it out, too many terms would become 0, causing a contradiction. If a state can be written as a tensor product, it's called *separable*; a state that cannot be written in this way is *entangled*.[4] This is all that entanglement is, the quantum-mechanical version of correlation. Entanglement should not be changed by local unitary transformations, though it may be destroyed by measurement (see below). A global unitary transformation, involving both qubits, could entangle or unentangle qubits.

In general, separability arises when we have a state space $\mathbb{C}^{AB} = \mathbb{C}^A \otimes \mathbb{C}^B$. This can get more interesting in infinite-dimensional Hilbert spaces, but most of the spaces we consider in this class will be finite-dimensional, so just $\mathbb{C}^N$ for some N. Some quantum systems appearing in quantum optics arise not as tensor products, but as symmetric products, which can cause people to get tangled up talking about entanglement.

How do we measure in the two-qubit system? It's simple if you present a state as $a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle$; the probability of $|00\rangle$ is $|a|^2$. Physically, though, this has surprising implications: the two qubits may be very far apart. If Alice has one and Bob has the other, then it's possible for Alice to only measure her qubit, which has state 0 with probability $\Pr[0] = |a|^2 + |b|^2$. If she observes 0, the two outcomes vanish: even before Bob can make a measurement, the state undergoes a *partial collapse* to $(1/\sqrt{|a|^2 + |b|^2})(a|0\rangle + b|1\rangle)$ (and something similar with c and d, if Alice sees 1). This may be generalized to systems of any dimension.

These statements are true for both separable and entangled qubits, but for separable qubits, they reduce to trivialities.

If Alice and Bob's qubits are in the Bell pair state, and Alice measures a 0, then she know that whenever Bob measures it, he will measure a 0 (and similarly, if she measures a 1, so must he). Bob's qubit "updates" instantaneously as soon as Alice measures it, no matter how far away they are. This is what famously unsettled Einstein, since it violates the relativistic principle that information cannot exceed the speed of light; this is so-called "spooky action at a distance." This doesn't seem useful for creating a faster-than-light telephone, since Alice has no control over the bit she sends. When you pick up a newspaper, that determines the headline on every copy of that newspaper, but that's not as spooky: there are other variables which explain the correlations without FTL travel. A similar result for the two qubits would be called a *local hidden-variable theory*, postulating a shared secret between the two qubits that explains the correlation.

It took about thirty years for the question to be formulated in this way and then to be answered.

[3] In practice, to save effort, these are simplified: $|0\rangle \otimes |0\rangle$ is denoted $|0\rangle|0\rangle$ or even $|00\rangle$.

[4] There are ways to quantify the amount of entanglement; this state, sometimes called the *Bell pair*, *EPR pair*, or *singlet state*, happens to be maximally entangled.

Theorem 1.4 (No-communication theorem)**.** *It's not possible to use entangled states for faster-than-light communication.*

So quantum mechanics does not break relativity. However, there is no hidden-variable theory, either: quantum mechanics is an intermediate point between the hidden-variable theory and true FTL communication. Yet if you wanted to simulate quantum mechanics in a classical universe, the simulation would need FTL communication.

Bell conducted an experiment that led to this conclusion, which was really an early phenomenon of a familiar concept in theoretical computer science, the two-prover game. There are three actors: Alice, Bob, and a referee. Alice and Bob cannot communicate, but the referee can send challenges to Alice and Bob and collect their responses. Alice and Bob are trying to cooperate, trying to get the referee to accept with the largest probability. They may plan a strategy in advance, but cannot communicate during the experiment, just like in a separated police interrogation.

In the modern reformulation of Bell's theorem, this game is called the *CHSH game*. The referee sends a random bit $x \in \{0,1\}$ to Alice and an independent random bit $y \in \{0,1\}$ to Bob. Alice sends back a random bit $a = a(x, r_a)$ and Bob sends back a random bit $b = b(y, r_b)$ (here, r_a and r_b are the sources of randomness for Alice and Bob, respectively). Alice and Bob win the game if $a + b = xy \pmod 2$.

Clearly this is not a game many people play for fun. Classically, Alice and Bob can win $3/4$ of the time by always responding 0 — and one can prove that, classically, there is no strategy that does better, a fact called *Bell's inequality*.

But if Alice and Bob shared a Bell pair of two qubits in advance, there is a way of correlating their measurements in this state such that their probability of winning is $\cos^2(\pi/8) \approx 0.85$. This is a lot more subtle than sending messages back and forth: by themselves, Alice and Bob don't notice anything special, since you need both of their answers. In this case, there can be no local hidden-variable theory, and entanglement is not just shared classical randomness.

The protocol takes a little time to explain, but Alice measures her qubit in a specific basis if she sees a 0, and in a different basis if she sees a 1, and Bob does something similar. One can show that the probability of winning is $\cos^2$ of the difference of their measurement angles, which can be as high as $\pi/8$. A second inequality, called *Tsirelson's inequality*, says that no matter how many qubits they share, Alice and Bob cannot do better.

Any theory with local hidden variables predicts that the success probability is at most $3/4$, but quantum mechanics doesn't, so quantum mechanics is not a hidden-variable theory. Bell never imagined his experiment to be actually carried out, but in the 1980s, people actually did this, and the universe is consistent with the predictions given by quantum mechanics. Most physicists weren't surprised: this was not the first experiment testing quantum mechanics, and it's passed all of them, and wasn't the last.

A neat slogan is that, like everything else, Bell's theorem comes down to interference influencing correlation. Bell's theorem, rather than getting into metaphysical questions, uses quantum entanglement to solve problems, and in this way anticipates the field of quantum communication.

Mixed states. Suppose Alice and Bob have qubits in a Bell pair. What state does Alice see? Naïvely, one might expect Alice to end up with $|+\rangle$, but if this were the case, then if she measured it in the $\{|+\rangle, |-\rangle\}$ basis, she should always get $|+\rangle$. So what does it mean for her to apply a Hadamard gate H to her qubit only? We take the tensor product $H \otimes I$: the Hadamard for Alice, and the identity for Bob. Often, the unitary operators we care about can be broken up into smaller components. One way of thinking about this: for all possible states of Bob's qubit ($|0\rangle$ and $|1\rangle$), Alice applies the Hadamard gate.

When Alice does this, the state looks like $(1/2)(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$: Alice observes $|0\rangle$ and $|1\rangle$ with equal probability. That's weird. The takeaway is that a rose (Bell pair) by any other name (basis) still smells as sweet (is still a Bell pair). Something new is happening: Alice's qubit is behaving more like a classical random bit than a quantum state.

To talk about a piece of an entangled system, one needs a more general description of states, called *mixed states*: these are probability distributions over different states. For example, as we just saw for the Bell pair, Alice's state is $1/2\ |0\rangle$ and $1/2\ |1\rangle$. To be clear: this is not a superposition, just a plain old random bit. This is a surprisingly classical form of uncertainly!

There's an important subtlety in mixed states, which is why people don't always think of them as probability distributions over pure states.[5] Specifically, there are different probability distributions that give rise to the same mixed state: Alice's mixed state is $1/2\ |0\rangle$ and $1/2\ |1\rangle$, but is indistinguishable from the mixed state $1/2\ |+\rangle$ and $1/2\ |-\rangle$: in any orthonormal basis, each of these produces each outcome half of the time. Writing out a mixed state as a distribution over pure states is redundant. Fortunately, there's a representation for mixed states that's not redundant, using what's called *density matrices*. These are a whole new (equivalent) way to view quantum mechanics itself, and is usually preferred by experimentalists.

Suppose I have a probability distribution of pure states $\{p_i, |\psi_i\rangle\}_{i=1}^n$; then, the corresponding density matrix is

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i|.$$

This is an $n \times n$ complex matrix.

For example, the density matrix for Alice's mixed state in the Bell pair is

$$\frac{1}{2}\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix} + \frac{1}{2}\begin{pmatrix}0\\1\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}.$$

You can compute that if we started with $(1/2, |+\rangle)$ and $(1/2, |-\rangle)$, we end up with the same matrix.

Generally, if we have an entangled system of the form $\sum \alpha_i |i\rangle|\psi_i\rangle$, then Bob's density matrix is

$$\rho_{\text{Bob}} = \sum_i |\alpha_i|^2 |\psi_i\rangle\langle\psi_i|.$$

This also eliminates global phase.

If U is a unitary matrix, then it acts on ρ by conjugation:

$$\rho \longmapsto \sum_i \rho_i U|\psi_i\rangle\langle\psi_i|U^\dagger = U\rho U^\dagger.$$

Another advantage of the density matrix is that measuring the mixed state in this basis just requires the diagonal entries: $\Pr[|i\rangle] = \rho_{ii}$.

That is, along the diagonal of a density matrix ρ, there's a probability distribution. Sometimes, that's all we have, and the density matrix is diagonal (including the Bell state). But density matrices may also have off-diagonal entries, e.g. the superposition $(1/\sqrt{2})(|0\rangle + |1\rangle)$. This is not the same, because in the $\{|+\rangle, |-\rangle\}$-basis, its outcome is always $|+\rangle$. Its density matrix is

$$\frac{1}{2}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}.$$

Experimentalists regard off-diagonal entries as the signature of quantum behavior. In practice, the diagonal has the largest terms, but the bigger the off-diagonal terms are, the better the experiment was (according to the experimentalists).

The no-communication theorem says that quantum entanglement still preserves locality. More precisely, if Alice and Bob have entangled quantum systems, there is no combination of unitary transformations and measurements that Alice can make to her system that changes Bob's density matrix, unless we condition on Alice's measurement outcomes. This is very similar to how measurements affect classical correlation. Since Bob's density matrix can be used to calculate every possible outcome of every possible measurement Bob can make, this theorem encompasses anything Alice and Bob can do.

Density matrices don't provide us any new physics. Given a density matrix $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, there's an equivalent pure state

$$\sum_i \sqrt{p_i}|i\rangle \otimes |\psi_i\rangle,$$

and from the perspective of the second observer, these look the same.

[5] A pure state is a degenerate mixed state, which assigns probability 1 to a single state.

Lecture 2.

The Quantum Toolbox: 9/12/16

Last time, we reviewed (or introduced) quantum mechanics, including state spaces, qubits, and measurements. We talked about entanglement, Bell's inequality, and the no-communication theorem, and what is and isn't counterintuitive about them. We also introduced density matrices, which provide an equivalent formulation for everything in quantum mechanics.

Today we'll cover a few remaining phenomena in quantum mechanics of one through three qubits, the distance between two quantum states, general notions of measure, and other ingredients that we'll need. Then, we'll introduce quantum circuits and build up to defining BQP, the complexity class bounded over polynomial-time quantum circuits.

These remaining phenomena will inform what we are and aren't allowed to do with qubits, which will come up again and again in quantum complexity theory. For example, we learned that measurement is destructive: it's modeled as an irreversible process. Wouldn't it be great if we could work around that? Reproducible measurements are a cornerstone of science, so it would be nice.

The no-cloning theorem. Specifically, what if we had a procedure that could copy states? We could start with a quantum state $|\psi\rangle$ and an *ancilla* (an extra qubit) $|0\rangle$, apply some unitary transformation and obtain $\{|\psi\rangle, |\psi\rangle\}$, so we could measure the first in one basis, and the second in another basis. However, this is not possible, thanks to the suggestively named no-cloning theorem.

Perhaps this is striking — in classical mechanics, it's very possible to copy information. This is one of the foundations of the Internet (as well as software and music piracy).

Suppose $|\psi\rangle = (\alpha|0\rangle + \beta|1\rangle)$, and we tensor that with the ancillary qubit $|0\rangle$. We want a unitary transformation

$$(\alpha|0\rangle + \beta|1\rangle)|0\rangle \longmapsto (\alpha|0\rangle + \beta|1\rangle)(\alpha|0\rangle + \beta|1\rangle) = \alpha^2|00\rangle + \alpha\beta|01\rangle + \beta\alpha|10\rangle + \beta^2|11\rangle.$$

So we want a 4×4 matrix sending $(\alpha, 0, \beta, 0) \mapsto (\alpha^2, \alpha\beta, \alpha\beta, \beta^2)$. This is not linear, so there's no such matrix. Thus, it's not possible to perfectly clone, and moreover this is a robust phenomenon: it's possible to prove theorems about approximate cloning, and only very weak approximations are allowed.

Another way to understand this is that unitary transformations preserve angles and inner products. Suppose $|\varphi\rangle$ and $|\psi\rangle$ are such that $|\langle\psi|\varphi\rangle| = c \in (0, 1)$: they're neither parallel nor perpendicular. If we were to clone this, we'd obtain $|\langle\psi^{\otimes 2} \mid U \mid \varphi^{\otimes 2}\rangle| = c^2 < c$ after acting by the unitary matrix U, which means this isn't actually unitary. In general, there are some irreversible operations that increase the inner product, but none decrease it.

If φ and ψ are orthogonal or parallel, then we can clone: this is essentially a reduction to the classical case, where information can be duplicated.

Monogamy of entanglement. Suppose in the classical world we have three bits that are correlated: if you look at any two, you know the value of the third (e.g. knowing they xor to 1). This is called a *promiscuous entanglement* (really).

Alternatively, consider three qubits in the three-qubit analogue of the Bell pair, called the *GHZ state*:

$$\frac{|000\rangle + |111\rangle}{\sqrt{2}}.$$

Give the first qubit to Alice, the second to Bob, and the third to Charlie. One day, Charlie isn't answering his calls, so how does this affect Alice and Bob? Let's compute the density matrix for Alice and Bob's qubits. If Charlie's qubit is a 0, both Alice and Bob have 0, and similarly for 1. Thus, the density matrix is

$$\begin{pmatrix} 1/2 & & & \\ & 0 & & \\ & & 0 & \\ & & & 1/2 \end{pmatrix},$$

which is a classical-like system, a coin flip. The idea is that Charlie might have measured his qubit, and the system acts like he did, summed over all possible measurements. All three are entangled, but no two are, like the Borromean rings of topology (see Figure 1).

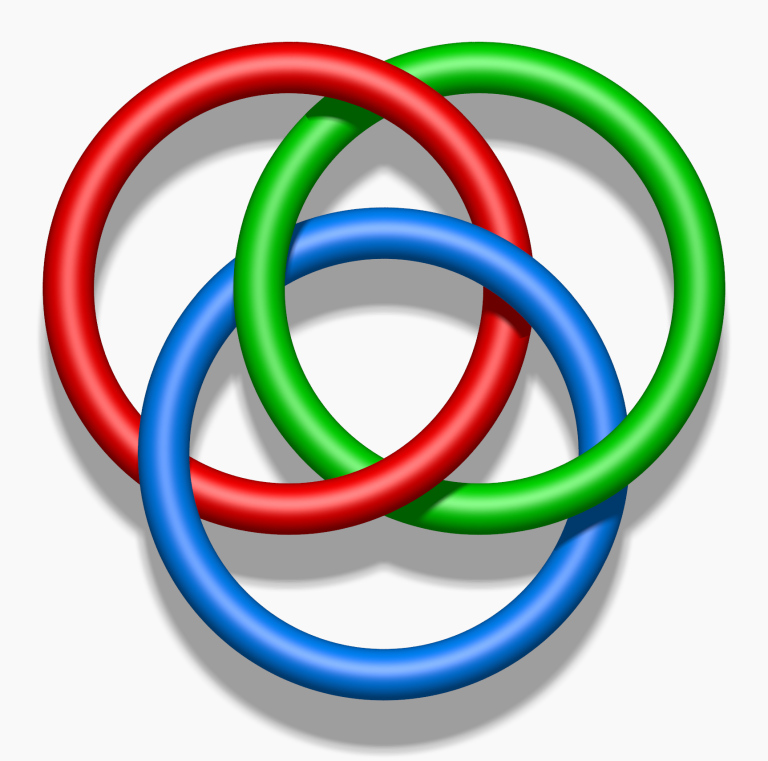

FIGURE 1. The Borromean rings, all three of which are linked, but no two of which are. Source: https://en.wikipedia.org/wiki/Borromean_rings.

This phenomenon is called *monogamy of entanglement*: if Alice and Bob are maximally entangled, Alice can't be maximally entangled with Charlie: quantum entanglement is "more jealous" than classical entanglement.[6]

Recall that a mixed state ρ_{AB} is separable if it's possible to write it as a probability distribution over (tensor) product states, and that entanglement is the absence of separability. This isn't a very efficient definition of entanglement, which may worry the complexity theorists in this course, and indeed: Gurvits proved in 2003 that deciding whether a state is separable is an NP-hard problem, by embedding the subset-sum problem into it. So unless P = NP, there's going to be no clean formula to detect entanglement.

Most reductions to NP are *Karp reduction*, in which a "yes" solution to problem A is translated to a "yes" solution to problem B in polynomial time, and similarly for "no" solutions. But there are also *Cook reductions*, in which one has an oracle for problem A and can make multiple queries to solve problem B. It was known for a while that there were some examples of NP problems which needed Cook reductions rather than Karp reductions, but Gurvits' proof was the first example for any real-world application.

Nobody knows the hardness of approximate separability (are these states almost separable?). There's some evidence that it's at least $O(n^{\sqrt{2}(\log n)})$, or else 3-SAT would be weirdly fast, but the problem is pretty wide open. So it's still quite hard to quickly tell whether states are entangled, and this frustrates a lot of experimentalists. There are some sufficient conditions, e.g. *entanglement witnesses* such as violating classical behavior such as the Bell inequality.

Quantifying entanglement and distances. The essential idea behind quantifying entanglement is called LOCC, for *local operations and classical communication*. This is a set of operations that don't increase entanglement: local unitary operators on Alice's or Bob's side, and classical communication. Any good measure of entanglement should be invariant under LOCCs.

[6] One often says that *maximal entanglement* means something that can be rotated into a Bell pair by a unitary matrix. This isn't so hard to define, but it's hard to come up with a good quantification of non-maximal entanglement. It gets more complicated for mixed states.

Definition 2.1. Suppose Alice and Bob have qubits in a state ρ_{AB}.

- The *entanglement of formation* $E_F(\rho_{AB})$ is the minimal number of Bell pairs needed to form ρ_{AB}.[7]
- The *distillable entanglement* $E_D(\rho_{AB})$ is the maximal number of Bell pairs that can be extracted from ρ_{AB} by LOCCs.

As an analogy, a stock has a buy price and a sell price.

Theorem 2.2. *For pure states, these two measures coincide, and also coincide with the Shannon entropy of* $|\alpha_i|^2$ *in the density matrix* $\sum \alpha_i |v_i\rangle|w_i\rangle$.

It's possible to use these to quantify monogamy of entanglement. For example, given a mixed state on three regions ρ_{ABC}, it's possible to show that

$$E_D(\rho_{AB}) + E_D(\rho_{BC}) \leq \log_2 \dim B.$$

For many things in this course, we'll need a way to measure distances. For pure states, the absolute value of the inner product is a good measurement, but in general, mixed states are more complicated. So we'll need to think about generalizations of probability distributions.

There are a few ways to measure distances of probability distributions, each of which has a quantum generalization. Probably the most useful measure is the *total variation distance*

$$\mathrm{TV}(\{p_i\},\{q_i\}) = \frac{1}{2}\|\{p_i\} - \{q_i\}\|_{L^1} = \frac{1}{2}\sum_i |p_i - q_i|.$$

This is a really nice measure: it defines a metric, satisfying the triangle inequality and all that, but it also has a nice statistical interpretation: it's the greatest possible difference between two outcomes of the same event, and there is an experiment producing this difference. This generalizes to quantum states.

Definition 2.3. Let ρ and σ be two mixed states, and let $\{\lambda_1, \dots, \lambda_n\}$ be the eigenvalues of $\rho - \sigma$. Then, their *trace distance* is

$$\|\rho - \sigma\|_{\mathrm{tr}} = \frac{1}{2}\sum |\lambda_i|.$$

Since (discrete) probability distributions are mixed states with diagonal matrices, this recovers the total variation distance.

Total variation distance 1 means that two distributions don't overlap, and are perfectly distinguishable. Similarly, two quantum states have trace distance 1 if they can be perfectly distinguished by a measurement: they live in orthogonal subspaces, in a sense.

Exercise 2.4. Show that the trace distance satisfies the triangle inequality, and therefore is actually a distance metric:

$$\|\sigma_1 - \sigma_3\|_{\mathrm{tr}} \leq \|\sigma_1 - \sigma_2\|_{\mathrm{tr}} + \|\sigma_2 - \sigma_3\|_{\mathrm{tr}}.$$

Moreover, since conjugation by a unitary matrix doesn't change eigenvalues, the trace distance doesn't depend on basis:

$$\|\rho - \sigma\|_{\mathrm{tr}} = \|U\rho U^\dagger - U\sigma U^\dagger\|_{\mathrm{tr}}.$$

Finally, we'll make use of the fact that if some measurement accepts ρ with probability p, it accepts σ with a probability in $[p - \delta, p + \delta]$, where $\delta = \|\rho - \sigma\|_{\mathrm{tr}}$.

Non-orthonormal bases. So far, all of our measurements have been in orthonormal basis. According to our rules, we don't actually know how to do anything else, but this was like pure vs. mixed states: we started with just pure states, and had to eventually draw out mixed states.

The idea is that if you entangle a qubit or two with some ancillas, measurements might take on more possible values, rather than just two. So we want to understand what can be measured when we're allowed to entangle an unlimited number of ancillary qubits.

One introduces the formalism of POVMs, or *positive operator-valued measures*, for this. Given a set of Hermitian, positive definite matrices E_i summing to the identity matrix, there is a procedure that returns i

[7] Technically, this is the limiting rate as we take more and more copies of this state, so isn't always an integer. The same is true for distillable entanglement.

with probability $\mathrm{Tr}(E_i\rho)$ for any state ρ, and every procedure defines a collection of matrices in this way. We won't prove this; see Nielsen-Chuang's book for more information.

For example, if ρ is an ordinary measurement (matrix) in the basis $|\psi_1\rangle, \ldots, |\psi_n\rangle$, then let $E_i = |\psi_i\rangle\langle\psi_i|$ (which, in this basis, has a 1 in position (i,i) and 0s everywhere else, projecting onto $|\psi_i\rangle$). Then,

$$\mathrm{Tr}(E_i\rho) = \mathrm{Tr}(|\psi_i\rangle\langle\psi_i|\rho) = \langle\psi_i \mid \rho \mid \psi_i\rangle,$$

which is how we defined measurement last lecture.

The POVM formalism is incomplete in that it doesn't specify the post-measurement state, and depending on implementation, the result isn't determined.

Superoperators. These various formalisms can be unified into *superoperators* $\rho \to \$(\rho)$, which can be thought of as "allowed operations on operators." For example, we could take $\$(\rho) = U\rho U^\dagger$ for some unitary operator U, or we could let $\$(\rho)$ zero out the non-diagonal entries. Another example maps every state to a particular state. Superoperators are deterministic operators from mixed states to mixed states, though the image may be a non-pure state (and thus probabilistic).

To formalize this, start with a set $\{E_1, \ldots, E_n\}$ of (not necessarily square) matrices such that

$$\sum_i E_i E_i^\dagger = I.$$

These define the superoperator

$$\$(\rho) = \sum_i E_i^\dagger \rho E_i.$$

One good exercise is to write down the matrices E_i for the three superoperators mentioned above.

There's a theorem that all measurements involving ancillary systems can be expressed using superoperators, unifying several of the formalisms we've defined so far.

$$\sim \cdot \sim$$

Let's combine these concepts into a lemma that we'll use many times. We've reiterated that measurement is a destructive process, but not all measurements are destructive; in a sense, measurements are only destructive when we don't have a basis vector near the state we're measuring. If you start with a state that's very close to $|0\rangle$, then almost all of the time, the state doesn't change by very much, so we haven't lost very much information. In quantum mechanics, this is sometimes referred to as the *information disturbance tradeoff*. The more randomness is generated, the more the state is disturbed.

Lemma 2.5 (Almost as good as new lemma or gentle measurement lemma)**.** *Let ρ be a mixed state and $M = \{E, I - E\}$ be a two-outcome POVM (so we accept ρ with probability $\mathrm{Tr}(E\rho)$), and suppose that $\Pr[M(\rho) \text{ accepts}] \geq 1 - \varepsilon$. Then, it is possible to implement M in such a way that the post-measurement state $\widetilde{\rho}$ satisfies $\|\widetilde{\rho} - \rho\|_{\mathrm{tr}} \leq \sqrt{\varepsilon}$.*

The intuition is that the trace distance is how much we've changed ρ by, but since the probability is high, the damage is small.

Proof. The first observation is that we can reduce to pure states, because any mixed state is a convex combination of projectors onto pure states:

$$\rho = \sum_i \rho_i |\psi_i\rangle\langle\psi_i|,$$

and the square root function is convex, so the sum of the square roots of the individual errors is at most the square root of the total error.

The second is that we only have to look at a two-dimensional subspace, since we've reduced to states of the form $|\psi\rangle \otimes |0\cdots0\rangle$. In this case, we're conjugating by a unitary matrix with (at least) $1 - \varepsilon$ and (at most) ε on the diagonals, so it has to look like

$$\begin{pmatrix} 1-\varepsilon & \sqrt{\varepsilon(1-\varepsilon)} \\ \sqrt{\varepsilon(1-\varepsilon)} & \varepsilon \end{pmatrix}.$$

After measuring, we simply have

$$\begin{pmatrix} 1-\varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix},$$

and you can calculate the trace distance to be $\sqrt{\varepsilon(1-\varepsilon)}$. ⊠

This is an important counterpart to the no-cloning theorem: measurement may destroy information, but if it accepts good values with near-certain probability, then the amount of damage done is very small. This is useful for quantum money, or copy-protected quantum software: you can prevent outright duplication, but observing that the certificate is valid doesn't destroy it.

In fact, multiple measurements are okay. The idea is that if there's at most a 1% chance that you'll fall off of cliffs and a 2% chance you'll be struck by lightning, you have at most a 3% chance of suffering both, no matter the correlations between the two events. The corresponding quantum version will be very useful.

Theorem 2.6 (Quantum union bound). *Suppose we apply POVMs $E_1, \dots, E_k$ in order to a mixed state ρ, and suppose that $\mathrm{Tr}(E_i\rho) \geq 1-\varepsilon$. Then,*

$$\Pr[E_1, \dots, E_k \text{ all accept}] \geq 1 - k\sqrt{\varepsilon}$$

and $\|\widetilde{\rho} - \rho\|_{\mathrm{tr}} \leq k\sqrt{\varepsilon}$, where $\widetilde{\rho}$ is the post-measurement state.

This does not follow immediately from Lemma 2.5.

Proof. We need to use the linearity of quantum mechanics. The triangle inequality tells us that

$$\|\rho - E_k \circ \cdots \circ E_1(\rho)\|_{\mathrm{tr}} \leq \|\rho - E_k(\rho)\|_{\mathrm{tr}} + \|E_k(\rho) - E_k \circ E_{k-1}(\rho)\|_{\mathrm{tr}} + \cdots + \|E_k \circ \cdots \circ E_2(\rho) - E_k \circ \cdots \circ E_1(\rho)\|_{\mathrm{tr}}.$$

Applying a superoperator can't decrease the trace distance, so

$$\leq \|\rho - E_k(\rho)\|_{\mathrm{tr}} + \|\rho - E_{k-1}(\rho)\|_{\mathrm{tr}} + \cdots + \|\rho - E_1(\rho)\|_{\mathrm{tr}}.$$

Applying Lemma 2.5,

$$\leq k\sqrt{\varepsilon}.$$ ⊠

This isn't a tight bound, and the bound has been improved to $\sqrt{k\varepsilon}$ in a recent paper.

One technical point is that we should pin down how to tell that all of the E_i accept, We can add ancillary qubits and entangle them to record whether a state accepts: if they all start with $|1_{E_i \text{ accepts}}\rangle$, they move by at most $\sqrt{\varepsilon}$ each, so we recover the bound we wanted.

$$\sim \cdot \sim$$

Quantum computing acts on systems of n qubits, where n may be large. Such a state may be represented by

$$|\psi\rangle = \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle,$$

where we normalize

$$\sum_{x \in \{0,1\}^n} |\alpha_x|^2 = 1.$$

This is kind of incredible: it tells us that to simulate 1000 particles, you need $2^{1000} \approx 10^{300}$ pieces of information, more pieces than all the subatomic particles in the universe. This growth in complexity is why problems like the Schrödinger equation are so difficult, and people earn Nobels for "small" advances.

This was turned on its head at first to provide more efficient simulations of quantum systems in physics or chemistry, and now has found additional applications. But we still have a little way to go in our formalization of quantum computers: we could also have been talking about a very classical notion of probability distributions on n bits, but nobody actually thinks of this as 2^n pieces of information. In quantum mechanics, though, these 2^n pieces of information are actually there in nature, thanks to interference. This suggests that creatively arranging interference can make for a computation that's a speedup from classical computation.

In some sense, this would all come from a $2^n \times 2^n$ unitary transformation, but not any unitary matrix (flip the n^{th} qubits iff the first $n-1$ define a Turing machine that solves the halting problem). This problem arises in classical mechanics too: the halting problem is a Boolean function,[8] but is not accessible to us.

For this reason, we look for Boolean functions that can be implemented with relatively few AND, OR, and NOT gates A general Boolean function on n bits can be implemented in $O(2^n/n))$ bits. This is due to

[8] A *Boolean function* is a function $f : \{0,1\}^n \to \{0,1\}$.

Shannon's counting argument: there are 2^{2^n} Boolean functions on n bits. Since the NAND gate is universal, we can ask how many ways there are to arrange T NAND gates. The first has $\binom{n}{2}$ possibilities to choose from, and the second can also ask for the output of the first, so it has $\binom{n+1}{2}$ possibilities. This can be approximated as

$$\binom{n}{2}\binom{n+1}{2}\cdots\binom{n+T-1}{2} \leq (n+T)^{2T}.$$

Thus, if we want to model all Boolean functions on n bits using at most T NAND gates, then $(n+T)^{2T} \geq 2^{2^n}$, i.e. $2T\log(n+T) \geq 2^n$, so $T = \Omega(2^n/n)$. This argument also shows that almost all functions require this many gates! But to know which ones do and don't at a large scale would provide a solution to whether P = NP: we know that only a small number of these functions are accessible or "nice."

Something very similar will happen in quantum computing: there will be a quantum circuit of a set of qubits, and an arrow of time. We can apply unitary operators as gates: applying the Hadamard operator H at time t to only bit 3 is akin to applying $I \otimes I \otimes H \otimes I$ to the whole system. This reflects the engineering behind all of this: quantum computers will be built out of small building blocks, so it's best to describe operations on those blocks.

We've already seen the Hadamard gate; let's discuss some others.

The *controlled NOT* gate is probably the next most important. If the first qubit is 1, it flips the second qubit; if the first qubit is 0, it doesn't. This can be written as

$$\text{CNOT} : |x, y\rangle \longmapsto |x, y \oplus x\rangle.$$

In truth-table form, $|00\rangle \mapsto |00\rangle$, $|01\rangle \mapsto |01\rangle$, $|10\rangle \mapsto |11\rangle$, and $|11\rangle \mapsto |10\rangle$. Like the Hadamard gate, this is its own inverse; in matrix form, it's

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

We can use this to entangle two states: starting with $|00\rangle$ and applying $H \otimes I$, then CNOT, the result is the Bell state, so you can entangle with two gates. Another weird fact is that, after a change of basis, a controlled NOT from x to y is equivalent to one from y to x.

This can shed light on some philosophy of quantum mechanics: if you want to measure a state $\alpha|0\rangle + \beta|1\rangle$, *decoherence theory* says we consider a system

$$(\alpha|0\rangle + \beta|1\rangle) \otimes |\text{Measurement}\rangle \otimes |\text{You}\rangle,$$

and there is some unitary transformation involving the atoms of the measuring device (and you!) with the outcome

$$(\alpha|0\rangle|\text{Measured } 0\rangle + \beta|1\rangle|\text{Measured } 1\rangle) \otimes |\text{You}\rangle.$$

When you look at it, this collapses to

$$\alpha|0\rangle|\text{Measured } 0\rangle|\text{You } 0\rangle + \beta|1\rangle|\text{Measured } 1\rangle|\text{You } 1\rangle.$$

This makes sense from a purely formal perspective, but actually thinking about what it *means*, or whether it means anything, is difficult and inconclusive. But as long as we talk about systems external to ourselves, this isn't as controversial.

Anyways, this provides an explanation for the CNOT gate: the target qubit is measuring the first qubit, just like entanglement. If you believe in the many-worlds perspective (which, well, is very metaphysical), then measurement is just a CNOT operation!

Definition 2.7. A gate set G is *universal* if we can use gates from G to implement any unitary operator on any number of qubits.

At some point, we'll worry about precision, but not yet. With no further conditions on the definition, these are necessarily infinite: there are uncountably many unitary operators, but a finite set can only generate countably many operators. Nonetheless, there is something nice to be said about these.

Theorem 2.8. *The set of* CNOT *and all* 1*-qubit gates is universal.*

We're not going to prove this.

Over a system of n qubits, this can be approximated by 4^n gates, which follows from a dimension-counting argument: the dimension of the manifold $\mathrm{U}(2^n)$ is 4^n. This means that we need systems of at least $T = 4^{n-1}$ gates to simulate all of these, since then $4T \geq 4^n$. Once again, the things that we can calculate efficiently are a vast subset of everything possible.

In real life, we cannot implement all of these perfectly: most are irrational (e.g. rotate by $1/e$), and some even are undecidable. For computational purposes, we only need to understand this to finite precision, thanks to the quantum union bound.

Definition 2.9. A gate set G is *approximately universal* if we can use gates from G to approximate any unitary transformation on any number of qubits to any desired precision.

Thanks to arbitrary precision, it doesn't matter what norm we use to evaluate this; they're all equivalent. This definition also allows finite gate sets to be approximately universal. We still don't completely understand approximately universal gate sets, but here are some nice facts.

Proposition 2.10. *For almost all* 1*-qubit gates* U,[9] $\mathrm{CNOT} + U$ *is (approximately) universal.*

Proposition 2.11. *For almost all* 2*-qubit gates* U, $\mathrm{CNOT} + U$ *is (approximately) universal.*

Example 2.12. The gate set

$$\left\{\mathrm{CNOT}, \begin{pmatrix} 3/5 & 4i/5 \\ 4/5 & -3i/5 \end{pmatrix}\right\}$$

is (approximately) universal.

The second matrix has to have is somewhere, so it can see[10] the unitary group rather than just the real orthogonal group; the real part of the above matrix, along with the CNOT gate, densely generate $\mathrm{O}(n)$.

The third most important gate is the *Toffoli gate*, sometimes called "controlled controlled NOT." It flips the third qubit iff the first and second qubits are 1, akin to a reversible AND. One can write an 8×8 matrix or summarize it as $|x, y, z\rangle \mapsto x, y, z \oplus xy$.

Proposition 2.13 (Shi, 2001)**.** *The gate set*

$$\left\{\mathrm{Toffoli}, H, \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}\right\}$$

is (approximately) universal.

It's also good to know which gate sets are not universal. These include gate sets that only act classically, e.g. CNOT plus a Toffoli gate, or a Toffoli gate and a phase: these won't put anything into superposition that wasn't already. There are more interesting examples; this next one is particularly not obvious.

Theorem 2.14 (Gottesmon-Knill)**.** *The gate set*

$$\left\{\mathrm{CNOT}, H, \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}\right\}$$

is not only non-universal, but generates a discrete subgroup of $\mathrm{U}(n)$*, and can be simulated by classical computation in polynomial time.*

This is surprising: these three gates alone do a lot of work, but to get a quantum speedup we need something else.

[9] This means if you pick a 1-qubit gate at random, the probability of it meeting this condition is 1.

[10] Pun intended?