Contents

Part I
Quantum Computing and Algorithms

1 Lecture 1

1.1 Motivating Quantum Computing

The classical unit of computation is a **bit**. How small can we shrink bits? Let's conduct a thought experiment. Let's suppose we could shrink them down to the size of a Hydrogen atom. The "state" of $|0\rangle$ being the ground state and $|1\rangle$ first excited state. However, electrons in general exist in superposition states! These states look like:

$$\{\alpha |0\rangle + \beta |1\rangle : \alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1\}$$

But it gets weirder. According to quantum theory, when conducting a measurement on such a state, we end up getting:

$$M = \begin{cases} 0 & \text{wp } |\alpha|^2 \\ 1 & \text{wp } |\beta|^2 \end{cases}$$

Furthermore, the act of measurement "collapses" the wavefunction to a state $|0\rangle$ or $|1\rangle$. Subsequent measurements will give that pure state deterministically.

Now, suppose we have a system of two such Hydrogen. There are now 4 basis states:

$$|\psi\rangle = \alpha_{00} |00\rangle + \alpha_{01} |01\rangle + \alpha_{10} |10\rangle + \alpha_{11} |11\rangle$$

Effective computation now comes from extrapolating to n such **qubits**. Now such a state would look like $\sum_{x \in \{0,1\}} \alpha_x |x\rangle$.

This is pretty profound. Classical computers were designed to use nature (through silicon) in order to work for humans. But with all this effective work that nature is doing behind the scenes, it seems that quantum computing is really the more powerful framework we should've asked for.

1.2 Measurement

Now suppose we do a "partial" measurement, e.g. only measuring the first bit. What will we get? It seems reasonable that the probability should be the sum of the probabilities of getting a 0 in the first qubit, e.g. we get a 0 w.p. $|\alpha_{00}|^2 + |\alpha_{01}|^2$. The state collapses, but it must be renormalized so the coefficients can still be probabilities! So the new state is actually

$$\frac{\alpha_{00} |00\rangle + \alpha_{01} |01\rangle}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}$$

Now suppose we are given a qubit in the state

$$|\psi\rangle = \frac{1}{\sqrt{2}} |0\rangle + \frac{e^{i\theta}}{\sqrt{2}} |1\rangle$$

How can we figure out θ (phase estimation)? Well if we measure this, we will get either 0 or 1 with probability $1/2$ each. This will tell us nothing about θ. It turns out this is only a special case of measurement.

To understand what general measurement is, we first go back to our state representation. What we really mean by a superposition is a linear combination of two vectors. We fix some basis $|0\rangle$ and $|1\rangle$, and a normalized state $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$ is a unit vector in a 2-dimensional complex vector space. Now we can think about a measurement in the following way:

Definition 1.1
A **measurement** of some state $|\phi\rangle$ in some basis $\mathcal{U}$ is a projection onto one of the basis vectors $|u\rangle$. The value of the measurement is: u with probability of the scalar projection squared, $\left|\frac{\langle u|\psi\rangle}{\langle\psi|\psi\rangle}\right|^2$.

So for example, let's stick to our 2-space and pick a new orthonormal basis $\{|u\rangle, |u^\perp\rangle\}$ and our state $|\psi\rangle$. Suppose $|\psi\rangle$ makes an angle of θ with the $|0\rangle$ axis and makes an angle of μ with the $|u\rangle$ axis. By a simple diagram, it's clear from ψ's projections that measurement in the standard basis yields 0 with probability $\cos^2\theta$ and in our new basis it yields u with probability $\cos^2\mu$.

Note 1.1
There is a bit of a subtlety here. We assumed that the amplitudes we are working with are real, but in general they can be complex. It turns out, all of quantum computing can be formalized with only real amplitudes, but it gets more messy when interfacing with physics. For now, we will assume real amplitudes only, but most results generalize to complex amplitudes.

Another common example of a basis is:

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

Measuring our original phase estimation in this new basis is exactly what we need! We just need to write it in the new basis to figure out the amplitudes:

$$\frac{1}{\sqrt{2}}|0\rangle + e^{i\theta}\frac{1}{\sqrt{2}}|1\rangle = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|+\rangle + \frac{1}{\sqrt{2}}|-\rangle\right) + e^{i\theta}\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|+\rangle - \frac{1}{\sqrt{2}}|-\rangle\right) \tag{1}$$

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

$$= \frac{1}{2}(1 + \cos\theta + i\sin\theta)|+\rangle + \ldots \tag{3}$$

$$\tag{4}$$

so we get $|+\rangle$ from the measurement with probability

$$\frac{1}{2}|1 + \cos\theta + i\sin\theta|^2 = \cos^2(\theta/2)$$

Now we can repeat the measurement (with other processed inputs) to get statistics and thus a good estimate on θ.

2 Lecture 2

2.1 Axioms of Quantum Mechanics

We list some axioms of Quantum Mechanics. Consider an electron with k energy levels, $|0\rangle, |1\rangle, \ldots, |k-1\rangle$.

Note 2.1 (Superposition Principle)
If there are k distinguishable (eigenstates) of a system, then the state of a system can be written as:

$$|\psi\rangle = \sum_{j=0}^{k-1} \alpha_j |j\rangle$$

where $\alpha_j \in \mathbb{C}$ and $\sum_j |\alpha_j|^2 = 1$.

This forms a Hilbert space, i.e. a Complex inner product space (but we will often think of all amplitudes as real). The $\{|j\rangle\}_{j=0}^{k-1}$ forms a basis for this state space. We can think of

$$|\psi\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{k-1} \end{pmatrix}, |0\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, |1\rangle = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots$$

For inner products, we use Dirac's Bra-Ket notation. As we have already seen, the "kets" are regular vectors and the "bras" $\langle\psi| = |\psi\rangle^\dagger$ are elements of the dual vector space (which can be thought of as conjugate transposes). This means:

$$\langle\psi| = |\psi\rangle^\dagger = \sum_j \left(\alpha_j |j\rangle\right)^\dagger = \sum_j \alpha_j^* \langle j|$$

where $(\cdot)^*$ is the complex conjugate.

Now define $|\phi\rangle = \sum_j \beta_j |j\rangle$. We can take inner products by using the following notation:

$$\langle\psi, \phi\rangle = \langle\psi|\phi\rangle = \left(\sum_i \alpha_i^* \langle i|\right)\left(\sum_j \beta_j |j\rangle\right) = \sum_{i,j} \alpha_i^* \beta_j \langle i|j\rangle = \sum_j \alpha_j^* \beta_j$$

Because $\langle i|j\rangle = 1$ if and only if $i = j$ (they form an orthonormal basis).

We generally use $k = 2$, call the Hilbert space generated $\mathcal{H}$. We typically think about chaining together (tensor-producting) this Hilbert space with itself n times. This is called a n-**qubit** state. A general state can then be written as:

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

with $\alpha_x \in \mathbb{C}$ and $\sum_x |\alpha_x|^2 = 1$.

Note 2.2 (Measurement Principle)
Pick an orthonormal basis $\mathcal{U} = |u_0\rangle, |u_1\rangle, \ldots, |u_{k-1}\rangle$. The outcome of a measurement is j with probability $\left|\langle u_j|\psi\rangle\right|^2$. In this process, the state is also perturbed and turned into the state $\left|u_j\right\rangle$

Look at last lecture for examples of measuring in different bases, with real amplitudes one can think about qubit states geometrically. The basis $\{|+\rangle, |-\rangle\}$ serves us well.

2.2 Bell Inequalities

Let us look more closely at combining two qubits, each with states $\alpha_0 |0\rangle + \alpha_1 |1\rangle$, $\beta_0 |0\rangle + \beta_1 |1\rangle$. We (tensor) product them together, producing a state:

$$|\psi\rangle = \alpha_0\beta_0 |00\rangle + \alpha_0\beta_1 |01\rangle + \alpha_1\beta_0 |10\rangle + \alpha_1\beta_1 |11\rangle$$

but most states are not a product of two states.

The Bell basis states are a common example of states which are **entangled**, e.g. cannot be written as "product states."

$$\left|\Phi^{\pm}\right\rangle = \frac{1}{\sqrt{2}} |00\rangle \pm \frac{1}{\sqrt{2}} |11\rangle, \left|\Psi^{\pm}\right\rangle = \frac{1}{\sqrt{2}} |01\rangle \pm \frac{1}{\sqrt{2}} |10\rangle$$

These four states form an orthonormal basis for two qubits.

Suppose your system was in the state Φ^+ and we did a partial measurement on the first qubit. Then with probability $1/2$ we collapse to $|00\rangle$ and with probability $1/2$ we collapse to $|11\rangle$. Note that we could achieve this in a classical sense too, with correlated ("glued") coin flips.

Furthermore, the Bell states are rotationally invariant.

Theorem 2.1
In any basis, we can write the Bell States as:

$$\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}} |00\rangle + \frac{1}{\sqrt{2}} |11\rangle = \frac{1}{\sqrt{2}} |vv\rangle + \frac{1}{\sqrt{2}} \left|v^\perp v^\perp\right\rangle$$

Let's prove this. Suppose $v = \alpha |0\rangle + \beta |1\rangle$. Then without loss of generality, we can write $v^\perp = -\beta^* |0\rangle + \alpha^* |1\rangle$. This means that:

$$\begin{aligned} |vv\rangle + \left|v^\perp v^\perp\right\rangle &= (\alpha |0\rangle + \beta |1\rangle)(\alpha |0\rangle + \beta |1\rangle) + (-\beta^* |0\rangle + \alpha^* |1\rangle)(-\beta^* |0\rangle + \alpha^* |1\rangle) \\ &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \end{aligned}$$

where some algebra is elided. Note that we could achieve this in a classical sense too, with correlated coin flips that are rotated.

To go beyond classical computation, we consider two qubit measurements. The first player measures in the standard basis and the second player measures in a new basis, $\{|v\rangle, |v^\perp\rangle\}$, rotated at an angle θ from the standard basis. The probability that these two measurements are unequal is $\sin^2\theta$ (for example, if the first measurement is 0, then the state is $|00\rangle$, so the component of $|v\rangle$ in the $|0\rangle$ direction is $\cos\theta$). However, classically, the probability that one observes a different outcome is in a sense proportional to θ.

So John Bell's experiment is as follows. Alice is given a uniformly random bit x and Bob is given a uniformly random bit y. They must each report back a bit a and b respectively. Alice and Bob "win" the game if $xy = a + b \pmod 2$.

They can play the game in two ways: either classically or quantumly. Classically, they cannot communicate (apart from maybe the "glued" coin). In the quantum setup, Alice and Bob share a Bell state. They both pick a uniformly random bit x_a, x_b respectively. If Alice is given 0, they measure their qubit in the standard basis, otherwise they measure it in a basis rotated by $\pi/4$. If Bob is given 1, they measure their qubit in a basis rotated by $\pi/8$, otherwise they measure in a basis rotated by $-\pi/8$. Call their measured bits a and b respectively.

We then mention the following two facts:

1. No classical strategy can win with probability $> 75\%$. A randomized strategy can do no better than a deterministic strategy since the opponent's strategy is known. The best deterministic strategy is to report $a = 0$ and $b = 0$ (or $a = 1$ and $b = 1$), because $xy = 0$ with probability 75% (if at least one of the bits is 0); trying to force the answer to be 1 will give you a lower probability of success. You can do no better. The glued coin doesn't help you either; the best it could do is give you a shared source of randomness.

2. In each of the 4 cases, the probability winning in a quantum setup is $\cos^2 \pi/8 \approx 85\%$. For example, take the case when x and y are both 0. Then they need to both measure a 1 or both measure a 0. The probability Alice measures a 0 is $1/2$ and then collapses the state to a $|00\rangle$. The probability that Bob then sees a 0 is $\cos^2 \frac{\pi}{8}$ because of the rotation, giving us $\frac{1}{2} \cos^2 \frac{\pi}{8}$. Likewise, the probability Alice measures a 1 is $1/2$ and then collapses the state to a $|11\rangle$. The probability that Bob then sees a 1 is $\cos^2 \frac{\pi}{8}$, so overall the probability is $2 \cdot \frac{1}{2} \cdot \cos^2 \frac{\pi}{8} = \cos^2 \frac{\pi}{8}$. The other cases are similar.

which clearly shows the quantum setup gives us something not present in the classical one.

3 Lecture 3

Recall the superposition and measurement principles from last lecture. They tell us that quantum states inhabit a Hilbert space $\text{span}\{|0\rangle, |1\rangle, \ldots, |k-1\rangle\}$ and we can "measure" in orthonormal basis in this Hilbert space, randomly projecting it onto a basis vector. These were two axioms of quantum mechanics.

3.1 Unitary Evolution

A third axiom of quantum information is the ability to apply a unitary transform. These are ubiquitous in linear algebra, but nonetheless we give a quantum-tuned introduction here.

For 1 qubit, we can think of a unitary transform as a "rigid-body rotation" (rotation/reflection), which preserves the orthogonality of vectors.

$$U = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

With our canonical representation of $|0\rangle = \begin{pmatrix} 1 & 0 \end{pmatrix}^T, |1\rangle = \begin{pmatrix} 1 & 1 \end{pmatrix}^T$, this transform can equivalently be stated as:

$$|0\rangle \mapsto a|0\rangle + b|1\rangle, |1\rangle \mapsto c|0\rangle + d|1\rangle$$

Define the adjoint of a matrix as its conjugate transpose, e.g.:

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

Now, we can interpret this in the 2-by-2 case as the following:

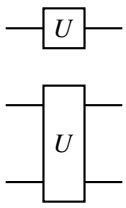
$$UU^\dagger = \begin{pmatrix} a^*a + b^*b & a^*c + b^*d \\ c^*a + d^*b & c^*c + d^*d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The last equality is only true if the columns of U are orthonormal, they are normalized (the top left and bottom right entries are just norms) and have inner product 0.

In general, we have the following definition:

Definition 3.1 (Unitary)
A transform $U \in \mathbb{C}^{n \times n}$ is unitary if and only if $UU^\dagger = U^\dagger U = I$, where I is the n-by-n identity.

Another name for these unitary transform are "quantum gates." We can draw such gates on one or two inputs as the following:



3.2 The Fundamental Quantum Gates

Some simple 1-qubit gates are the following:

Definition 3.2

1. The identity gate, which takes a state and does nothing:

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

2. The rotation gate, which rotates a state by θ:

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

3. The inversion (NOT) gate, which flips a bit:

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

   i.e., $a|0\rangle + b|1\rangle \mapsto b|0\rangle + a|1\rangle$.

4. The phase flip gate, which flips the phase of the second bit:

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

   i.e., $a|0\rangle + b|1\rangle \mapsto a|0\rangle - b|1\rangle$

5. The Hadamard gate, which converts to the $\{|+\rangle, |-\rangle\}$ basis.

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

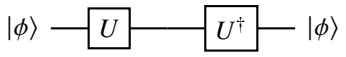
   i.e. $a|0\rangle + b|1\rangle \mapsto a|+\rangle + b|-\rangle$. One can view the Hadamard gate as a reflection over the line $\theta = \pi/8$. But hang on, we only allowed rotations, so what gives? It turns out the Hadamard gate is a rotation in $\mathbb{C}^2$, but not in $\mathbb{R}^2$!

Note that $X^2 = Z^2 = H^2 = I$, so they are involutions (and thus their own inverses). Furthermore X and Z are the same under a change of basis (note that $Z|+\rangle = |-\rangle$ and $Z|-\rangle = |+\rangle$).

$$X = HZH, Z = HXH$$

If you recall the Pauli spin matrices, you may remember X, Z as two of them; however, they are not expressive enough to correspond to all unitaries! Using H is computationally more interesting and is helpful for our analysis.

Let us make a bit of a silly circuit:



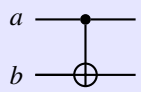
We applied a gate and then applied its adjoint, which is its inverse since it is unitary. Thus, we can always "undo/uncompute" quantum circuits (before measurement).

Now let us define a two-qubit gate,

Definition 3.3 (Controlled NOT)
The CNOT gate is:

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

To draw a CNOT gate, we draw it as the following:

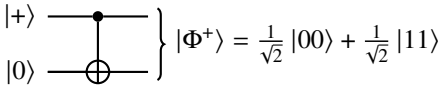


a is called the "control bit" and b is called the "target bit."

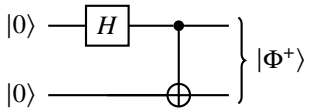
If a and b are pure bits, then we have the following truth table (the first bit controls whether a NOT gate is active, i.e. you XOR the two bits):

a	b	a_o	b_o
0	0	0	0
0	1	0	1
1	0	1	1
1	1	1	0

What if I did the following? I will input $|+\rangle|0\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|10\rangle$.



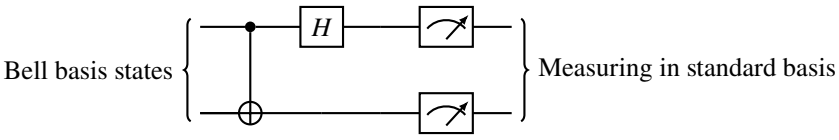
We created a Bell-state. Now to make it from two $|0\rangle$'s, we can add a Hadamard:



Now consider applying this circuit to any two-qubit state. We can do something very similar (the reader can verify the details):

Input	Output
$\|00\rangle$	$\|\Phi^+\rangle$
$\|01\rangle$	$\|\Psi^+\rangle$
$\|10\rangle$	$\|\Phi^-\rangle$
$\|11\rangle$	$\|\Psi^-\rangle$

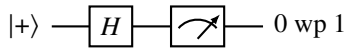
Now, consider turning the circuit backwards. Since both gates are their own inverses (CNOT is made up of a block diagonal of involutions so it is also an involution), this just inverts the circuit. Let us add a measurement apparatus:



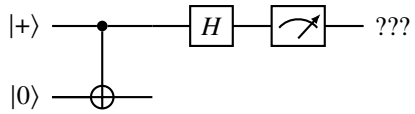
Now, we can measure in the Bell-state basis with *just* using our module for measuring the standard basis. This means without loss of generality we can measure in any basis.

3.3 Intuition for Entanglement

We know that:



But, now what if we entangle the state with a CNOT?



Intuitively, we expected the measurement controlled bit to not get changed, so we should expect the same result as above.

But let's analyze it formally. After the CNOT, the state is $|\Phi^+\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$. Then, the Hadamard doesn't act on the second bit, but the first bit is split, e.g. the final state is:

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|10\rangle\right) + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|01\rangle - \frac{1}{\sqrt{2}}|11\rangle\right) = \frac{1}{2}|00\rangle + \frac{1}{2}|01\rangle + \frac{1}{2}|10\rangle - \frac{1}{2}|11\rangle$$

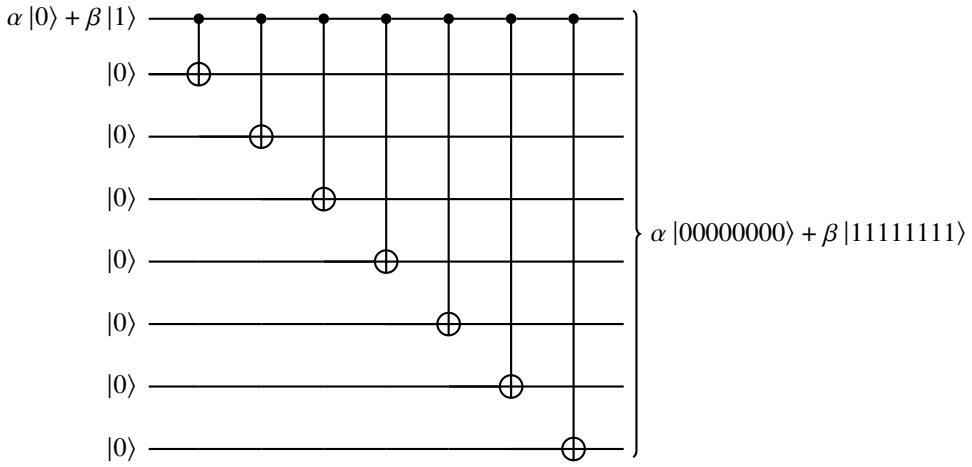
Measuring only the first qubit actually makes it so that we actually have a $\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2}$ chance of measuring a 0.

What happened here? In the first case, there is a cancellation of the probability amplitudes:

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

But in the second case, we have an entanglement which stops this cancellation (the product states cannot just be cancelled).

This gives us a view into what measurement really is. What if we entangle a bunch of bits:



At some macro point, nature cannot support such a large entangled state and collapses it probabilistically into one of the two basis states. This is how measurement is done in practice.

4 Lecture 4

4.1 The Tensor Product

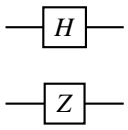
Remember that we "multiplied" two states $|0\rangle |0\rangle = |00\rangle$. This combined two 2-state systems into a 4-state system. But what is this mystical multiplication? The answer is the **tensor product**, denoted by the symbol $\otimes$. Really, it would be proper to say $|0\rangle \otimes |0\rangle = |00\rangle$, in which we "call" the tensor product this nickname. We also saw that we could identify $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ with vectors in $\mathbb{C}^4$, in this sense, we say $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^4$, i.e. the spaces are isomorphic.

In a general setting, letting $\mathcal{H}_1$ be a system with k levels and $\mathcal{H}_2$ be a system with ℓ levels, we say

$$\mathcal{H}_1 \otimes \mathcal{H}_2 \text{ is the vector space spanned by } \{|i\rangle |j\rangle\}_{0\le i\le k, 0\le j\le \ell}$$

which is a vector space with dimension $k \cdot \ell$.

We can also talk about the tensor product of operators. When you tensor product two operators, they act on their spaces separately. This is a circuit representation of $H \otimes Z$:



In this circuit, $|00\rangle \mapsto H|0\rangle \otimes Z|0\rangle = |+\rangle |0\rangle$, keeping the tensor products separate. On the other hand, note that the operator can be equivalently written as an element of $\mathbb{C}^{4\times 4}$.

$$H\otimes Z = \begin{pmatrix} |0\cdot\rangle \to |0\cdot\rangle & |1\cdot\rangle \to |0\cdot\rangle \\ |0\cdot\rangle \to |1\cdot\rangle & |1\cdot\rangle \to |1\cdot\rangle \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} & \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} & -\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$$

The first matrix represents the parts of the subspace that the matrix acts on. The notation $|b\cdot\rangle$ corresponds to the subspace first bit being b and the second bit being anything.

In addition, we will state one more fact about tensor products, which we will not prove:

Theorem 4.1
The inner product multiplies under the tensor product.

$$(\langle u| \otimes \langle v|)(|x\rangle \otimes |y\rangle) = \langle u|x\rangle \langle v|y\rangle$$

4.2 No Cloning Theorem

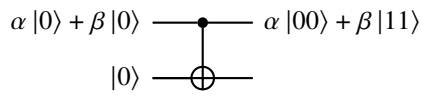
The no cloning theorem states that you cannot create a copy of a qubit. Formally,

Theorem 4.2 (No Cloning)
There is no unitary transform (quantum circuit) U such that

$$U|0\rangle |\psi\rangle = |\psi\rangle |\psi\rangle$$

for all possible states that $|\psi\rangle$ could take on.

But hang on, didn't we see a copying circuit before? Recall the following circuit:



It sent $|0\rangle \mapsto |00\rangle$ and $|1\rangle \mapsto |11\rangle$, which uniquely defines the transformation. But it doesn't work on a general state! By linearity note that:

$$(\alpha|0\rangle + \beta|1\rangle)|0\rangle = \alpha|00\rangle + \beta|10\rangle \mapsto \alpha|00\rangle + \beta|11\rangle \neq (\alpha|0\rangle + \beta|1\rangle) \otimes (\alpha|0\rangle + \beta|1\rangle)$$

This doesn't work.

To prove the no-cloning theorem another way, we then suppose for the sake of contradction that there exists a unitary U that can clone. Then for two vectors:

$$U|\phi\rangle|0\rangle = |\phi\rangle|\phi\rangle, U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle$$

But a unitary cannot change the angle between two vectors. To begin with and end with, the inner product was:

$$\langle\psi|\phi\rangle\langle 0|0\rangle = \langle\psi|\phi\rangle\langle\psi|\phi\rangle \implies \langle\psi|\phi\rangle = 0, 1$$

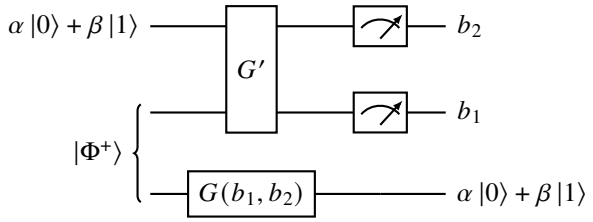
Any unitary can only clone only orthogonal states or the same state. In general, you can only clone some orthonormal basis of your choosing, like $\{|0\rangle, |1\rangle\}$.

4.3 Superdense Coding

Let's say Alice and Bob share a Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$. Alice has two bits of information $b_1, b_2 \in \{0, 1\}$ that she wants to send to Bob. Classically she can send this information in exactly two bits, no more, but with a quantum state, Alice can do it with just one qubit. Based on the bits, she turns the state into one of the Bell states. Alice then sends Bob her qubit. Now Bob can just measure the qubit in the Bell basis. But can we do better than a rate of 2 classical bits for one qubit? It turns out we can't.

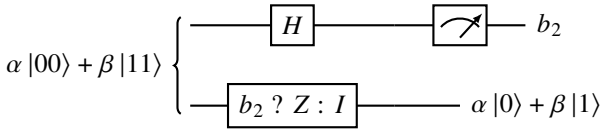
4.4 Quantum Teleportation

Professor Vazirani assures us that Alice has been working hard in her lab and made a state $\alpha|0\rangle + \beta|1\rangle$ that takes a LONG time to make. Unfortunately, she does not have access to an apparatus needed to process this qubit into cool things. Fortunately, she shares a Bell state with Bob and can exchange (classical) information with him, who does have such an apparatus. As we will see, the following circuit allows Alice to share her state with Bob:



where b_1 tells you whether to do a bit flip and b_2 tells you whether to do a phase flip in G.

Let's consider a simpler problem, Alice and Bob have two entangled bits $\alpha|00\rangle + \beta|11\rangle$ and they want Bob to create $\alpha|0\rangle + \beta|1\rangle$. They can use the following circuit to solve this problem:

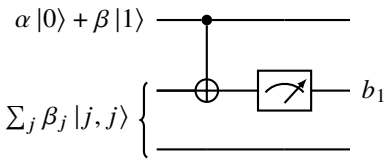


The state becomes:

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

Upon measurement, we get 0 and the state is $\alpha |0\rangle + \beta |1\rangle$, or we get 1 and the state is $\alpha |0\rangle - \beta |1\rangle$. She tells Bob this one bit b_2. If it were 1, Alice would tell Bob to apply a phase flip Z to fix the quantum state.

Now, to address the original problem, we can reduce to this case by just getting $\alpha |00\rangle + \beta |11\rangle$. The following subcircuit takes the two states and gives a 0 if the state is $\alpha |00\rangle + \beta |11\rangle$ and a 1 if the state is $\alpha |01\rangle + \beta |10\rangle$. In the second case, a bit flip on the second qubit gets us our target state.



Let's prove this claim, using an index for brevity, where the initial state of the (technically three) qubits is: $\sum_{i,j} \alpha_i |i\rangle \otimes |j, j\rangle$. Under CNOT:

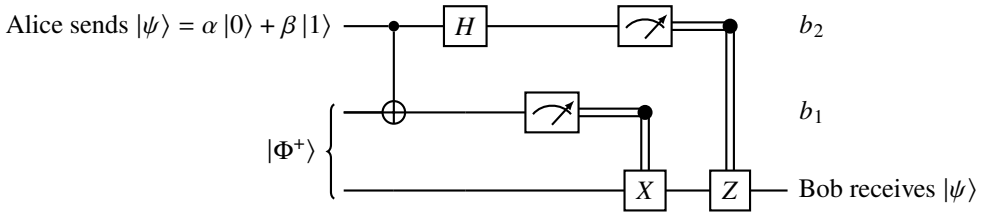
$$\text{CNOT} \sum_{i,j} \alpha_i |i\rangle \otimes |j, j\rangle = \sum_{i,j} \alpha_i |i, i \oplus j, j\rangle$$

But then after measurement of ℓ in qubit 1, we have $i \oplus j = \ell \implies j = i \oplus \ell$ is the only term that survives:

$$\sum_i \alpha_i |i, i \oplus \ell\rangle$$

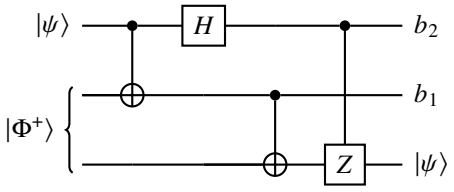
But we want to end up with $\sum_i \alpha_i |i, i\rangle$. So we only have to do a bit flip on the third bit if $\ell = 1$.

Combining all these subcircuits together gives us the following circuit



which allows Bob to successfully receive Alice's qubit. Using this circuit, we can teleport Alice's qubit anywhere!

Note that due to the principle of deferred measurement (covered in the next lecture), this circuit is equivalent to



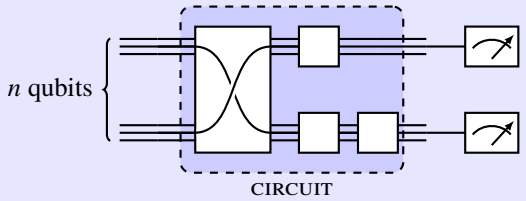
5 Lecture 5

The midterm is coming next Tuesday. It should serve as a checkpoint for what we've learned so far. The material in scope is the first 4 lectures plus the Hadamard transform we will define today.

5.1 Quantum Circuits and Quantum Algorithms

We will now begin quantum computer science in an algorithmic sense.

Definition 5.1 (Quantum Circuit)
A quantum circuit on n qubits is a collection of m gates (unitary transforms), with measurement at the end.



The depth d of a quantum circuit is the maximum amount of gates any one qubit is passed through. At the end of it, we measure in the standard basis in some qubits.

One way to think about a quantum circuit is "rotating" a state a lot of times and then measuring in the standard basis, or we could think about a quantum circuit as measuring in some "rotated" basis. In the quantum world, "programming" is really just designing such a quantum circuit. In a typical case, we want $m = O(\text{poly}(n))$, i.e. on the order of some polynomial of the number of qubits.

Suppose we had a (classical) circuit that could compute a Boolean function $f : \{0,1\}^n \to \{0,1\}^k$:

$$x \equiv \boxed{C_f} \equiv f(x)$$

How can we create such a circuit in the quantum world? Well first, we'd like:

$$|x\rangle \equiv \boxed{U_f} \equiv |f(x)\rangle$$

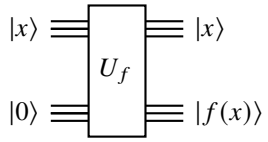
So by linearity,

$$U_f\left(\sum_x \alpha_x |x\rangle\right) \to \sum_x \alpha_x |f(x)\rangle$$

Remember that in our analysis here, since f is a Boolean function, we assume that x is a classical state, i.e. $|x_1, x_2, \ldots, x_n\rangle$ with $x_i \in \{0,1\}$. Furthermore, such a circuit is always invertible (since it's always unitary!):

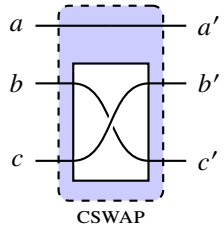
$$|f(x)\rangle \equiv \boxed{U_f^\dagger} \equiv |x\rangle$$

But this is clearly a problem! Suppose $f(x)$ was not injective, i.e. there existed two inputs $x \neq x'$ such that $f(x) = f(x')$. Clearly you cannot "go backwards." This makes us think that maybe all Boolean functions are not representable as quantum circuits. Thus we can only do it for f that are bijections! If we put the input included with the output, then we can go ahead and do this. Note that this does not break no-cloning because we are only cloning a pure binary state $x \in \{0,1\}^n$.



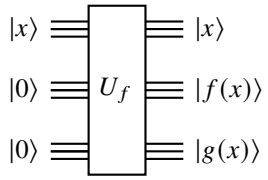
But note that all gates we use must also be reversible in the same. Recall that any classical circuit can be produced from AND and NOT gates. The NOT gate is invertible, and has quantum analogue X. But the AND gate is obviously not reversible.

However, it's very easy to make a reversible AND gate, we just use the same trick: just keep the inputs at the output. However, in practice, the more elegant way to do this is the CSWAP gate.



If $a = 0$, then $b' = b$ and $c' = c$, but if $a = 1$, $b' = c$ and $c' = b$. We claim this implements and AND gate if we set $c = 0$. Then, $c' = 0$ if $a = 0$ and $c' = b$ if $a = 1$, i.e. $c' = a \wedge b$. Furthermore, we can even implement a fanout, to "clone" an input. One can see through a calculation that if $b = 1, c = 0$, then $a' = a, b' = \bar{a}, c = a$.

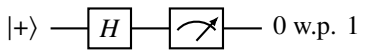
Note that we do not want a lot of these bits to actual AND them, they are useless to us. So we can add some "garbage" bits to the output:



But this is a disaster! Why? We shall explain soon. First, let's discuss how to fix it. To fix this, we can just copy $|f(x)\rangle$ (using a bunch of CNOTs), and then run $U_f^\dagger$ on everything else (including the original $|f(x)\rangle$). This eliminates the garbage $g(x)$ (the output of the $U_f^\dagger$ is the just input padded with 0's), while preserving $|f(x)\rangle$.

5.2 Principle of Deferred Measurement

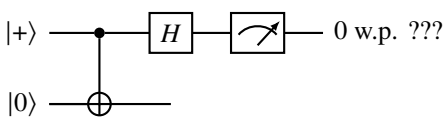
Let's go back to a basic quantum circuit.



This is because:

$$\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \overset{H}{\mapsto} \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle\right) + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle\right) = |0\rangle$$

Now, let's go to a more complicated circuit:



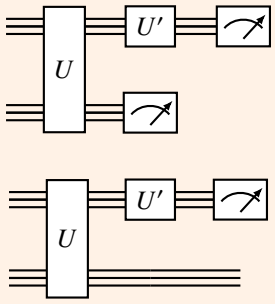
The output of the CNOT is fed into the H:

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle \overset{H}{\mapsto} \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|10\rangle\right) + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|01\rangle - \frac{1}{\sqrt{2}}|11\rangle\right)$$
$$= \frac{1}{2}|00\rangle + \frac{1}{2}|10\rangle + \frac{1}{2}|01\rangle - \frac{1}{2}|11\rangle$$

Now we don't have the interference that happened before where the $|1\rangle$ cancelled out. So now, upon measurement we see 0 and 1 with equal probability.

Let's generalize this.

Theorem 5.1 (Principle of Deferred Measurement)
After qubits stop interacting, the state is the same as if you measured them instead. For instance, the following two circuits will conduct equivalent measurements on the beginning bits:



So now, going back to the garbage bits, we now see why we cannot just "throw them away" or measure them. This will mean you can no longer use those bits for computation! As quantum information scientists, we have to keep our workspace "clean," so to speak.

5.3 Hadamard Transform

Let's send a n-bit classical bit state $|u\rangle$ into a bunch of Hadamards tensored together. Call $H^{\otimes n} = \underbrace{H \otimes H \otimes \cdots \otimes H}_{n \text{ times}}$.

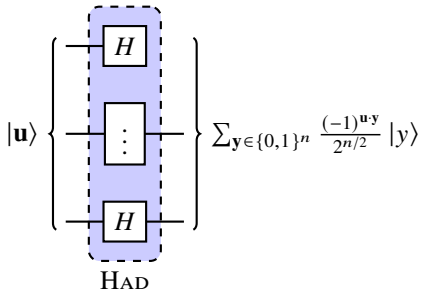
At the level of a singular qubit, we have:

$$H|u_1\rangle = \sum_{y \in \{0,1\}} \frac{(-1)^{u_1 \cdot y}}{\sqrt{2}}|y\rangle$$

which can clearly be seen by the definition (we only have a negative sign if both u and y are 1). Similarly, by nearly "squaring" the above expression, we have:

$$(H \otimes H)(|u_1, u_2\rangle) = \sum_{\mathbf{y} \in \{0,1\}^2} \frac{(-1)^{u_1 y_1}(-1)^{u_2 y_2}}{\left(\sqrt{2}\right)^2}|y\rangle$$

Thus, we can write:



where $\mathbf{u} \cdot \mathbf{y} = \sum_{i=1}^{n} u_i y_i \pmod 2$ is the dot product of the two bitstrings.

Lastly, one can compute the tensor product using the formula from last lecture:

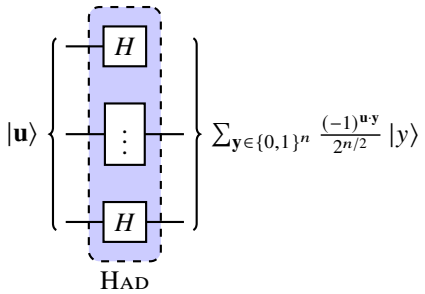
$$H \otimes H = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

One can continue this with higher order products in a similar fashion.

6 Lecture 6

6.1 Hadamard Transform (Continued)

Recall the Hadamard Transform from last lecture, made up of a bunch of Hadamard gates tensored together.



Recall that such a $|u\rangle \in \{0,1\}^n$ is a computational basis vector. So in the standard basis, it has a 1 in some position, and a 0 everywhere else. We will call this position the uth position. Thus, this picks out the uth column of $H^{\otimes n}$. This means the uth column is just $\frac{(-1)^{\mathbf{u}\cdot\mathbf{y}}}{2^{n/2}}$ for all possible y. Therefore, we have

$$[H^{\otimes n}]_{y,u} = \frac{(-1)^{\mathbf{u}\cdot\mathbf{y}}}{2^{n/2}}.$$

Note that these are just bitstrings, but when we use them as indices, we convert them to their decimal counterparts (and use 0-indexing).

One can view this as a Discrete Fourier Transform over $\mathbb{Z}_2^n$, which means it will be very nice for use later. Let's revisit the special case of 2x2.

$$H^{\otimes 2} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

Let's think about how to act on one of our favorite states, $|\Phi^+\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$. The $|00\rangle$ maps to the first column, $\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$ and the $|11\rangle$ maps to the last column, $\frac{1}{2}(|00\rangle - |01\rangle - |10\rangle + |11\rangle)$. By linearity, this means:

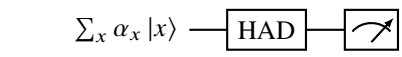
$$H^{\otimes 2}\left|\Phi^+\right\rangle = \frac{1}{2\sqrt{2}}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) + \frac{1}{2\sqrt{2}}(|00\rangle - |01\rangle - |10\rangle + |11\rangle) = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$$

That means this Bell state is an eigenvector of the Hadamard transform.

With some similar algebra, it turns out another Bell state can be mapped as:

$$H^{\otimes 2}\left(\frac{1}{\sqrt{2}}|01\rangle + \frac{1}{\sqrt{2}}|10\rangle\right) = \frac{1}{\sqrt{2}}|00\rangle - \frac{1}{\sqrt{2}}|11\rangle$$

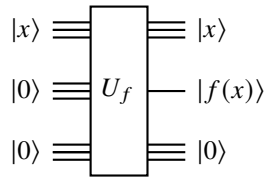
We can use the Hadamard to do something interesting:



Suppose the state before measurement is $\sum_y \beta_y |y\rangle$. Then measuring in the standard basis gives y with probability $|\beta_y|^2$. The question is, what are some interesting input states we can put into the Hadamard? Thinking about this question lays the groundwork for quantum algorithms.

6.2 Building Blocks for Quantum Algorithms

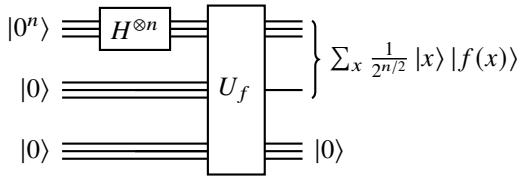
How do we program a quantum computer? Let's again try to emulate a classical computer. We saw last lecture, we can emulate a classical circuit C_f using U_f:



Now, suppose given a function $f : \{0,1\}^n \to \{0,1\}$, we want to make a circuit that produces:

$$\sum_x \frac{(-1)^{f(x)}}{2^{n/2}} |x\rangle$$

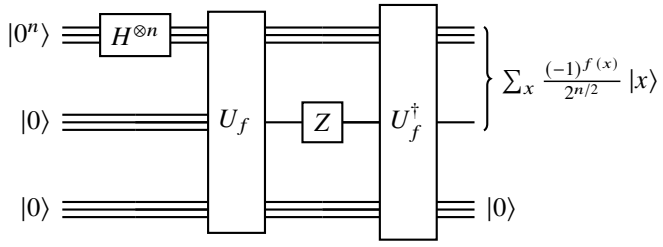
Our first attempt is:



We're close, but we need to somehow obtain the $(-1)^{f(x)}$. We can do this by applying a phase flip Z to the middle $|f(x)\rangle$ bit, which negates every term where $f(x) = 1$, i.e., the desired output becomes:

$$\sum_x \frac{(-1)^{f(x)}}{2^{n/2}} |x\rangle |f(x)\rangle$$

Now, to erase $|f(x)\rangle$, we can do the usual trick of running $U_f^\dagger$ on everything to undo the transform. Then, we measure to yield our intended answer (as $U_f^\dagger |x\rangle |f(x)\rangle = |x\rangle$). The circuit that puts this all together is depicted below:



6.3 Motivation for Quantum Algorithms

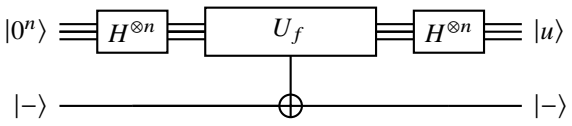
Before building up quantum algorithms, we thought about the Extended Church-Turing Thesis, which roughly stated that any "reasonable" model of computation is polynomial-time simulable on a (probabilistic) Turing Machine. Clearly, simulating n qubits with a classical computer requires tracking 2^n amplitudes, which is an exponential-time process. However, quantum computers violate this thesis and early quantum algorithms showcased this fact. Recently, Google performed a "Quantum supremacy" experiment, where they argued there was a **experimental** speedup from a classical computer running the same task.

6.4 Bernstein-Vazirani Algorithm

Suppose there is a secret function $f(x) = u \cdot x$ where we don't know u, but we have oracle access to it. The algorithm to solve this can be stated simply.

1. Create a phase state $\sum_x \frac{(-1)^{f(x)}}{2^{n/2}} |x\rangle = \sum_x \frac{(-1)^{u \cdot x}}{2^{n/2}} |x\rangle$
2. Feed it through the Hadamard Transform (it is its own inverse), and then measure. This will yield u.

The quantum circuit that performs this algorithm is depicted as follows:



where the $|-\rangle$ control performs the same operation as the Z gate described previously (i.e. generates the $(-1)^{f(x)}$).

But how would you figure this out classically? To uniquely find f, you'd need to put in a full basis of x, which necessitates at least n queries. On the other hand, in the quantum setting, Bernstein-Vazirani requires only two query accesses (assuming you can invert f). This is a substantial speedup.

You can generalize this algorithm with one called **Recursive Fourier Sampling (RFS)**. This is a similar algorithm that works on multi-dimensional vectors, allowing you to compute gradients (see that here we computed the gradient of f). For the problem that RFS solves, the comparison of runtimes between classical algorithms and quantum algorithms is shown below:

- Classical Algorithms satisfy the recursion $T(n) > nT\left(\frac{n}{2}\right) + n$. Solving this recurrence yields a runtime of

  $$T(n) = \Omega(n^{\log n}),$$

  which is *superpolynomial*.
- Our Quantum Algorithm (i.e. RFS) satisfies the recursion $T(n) = 2T\left(\frac{n}{2}\right) + O(n)$. Solving this recurrence yields a runtime of

  $$T(n) = O(n \log n),$$

  which is *polynomial*.

Here, we can see the power of quantum algorithms for computational optimization.

7 Lecture 7

7.1 Simon's Algorithm

We discuss Simon's algorithm: another quantum protocol that solves a toy problem. The setup is as follows. Suppose there is a black-box $f : \{0,1\}^n \to S \subseteq \{0,1\}^n$ that is two to one in a specific way: for a fixed string s, $f(x) = f(x \oplus s)$ for all $x \in \{0,1\}^n$, where the $\oplus$ is vector addition mod 2. Our challenge is to find s.

Classically, there are 2^n possibilities for s, so if we query $x \neq x'$ and get $f(x) \neq f(x')$ we can only cross out one possibility for $s \neq x \oplus x'$. So in the worst case, it takes 2^n tries. On average, due to the Birthday paradox, the runtime is actually the square root of this, $2^{n/2}$. However, the runtime is still exponential. Let's switch over to the quantum world to see if we can speed things up.

Here is a quantum algorithm that solves the same problem (with some randomness tossed in as well). Let $a \in_R A$ denote that a is uniformly chosen from A. Here is a brief roadmap of the steps we will take.

1. Set up superposition:
$$r \in_R \{0,1\}^n, \quad |\psi\rangle = \frac{1}{\sqrt{2}}|r\rangle + \frac{1}{\sqrt{2}}|r \oplus s\rangle$$
2. Perform Fourier sampling on $|\psi\rangle$, i.e. apply $H^{\otimes n}$, and measure. We claim that this yields uniformly random $a \in \{0,1\}^n$ such that $a \cdot s = 0$.
3. Wait until we get $n-1$ equations. Solve the linear equations for s. This takes time polynomial in n, classically.
4. Check if the solution you got to the system is correct by checking if $f(0) = f(s)$. Repeat the algorithm if not.

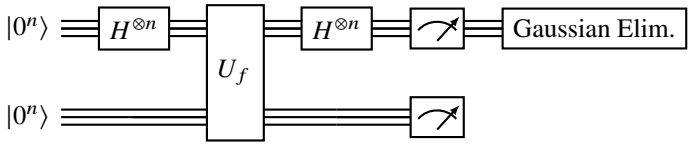
In the last step, the reason we use $n-1$ equations is that we will always have at least two solutions (0 and s), so no n equations can all be independent.

Let's figure out how to set up a suitable superposition for step 1. We first apply a Hadamard to $|0\rangle$ to make $\frac{1}{\sqrt{2^n}}\sum_x |x\rangle$. Then, if we apply the function f, we recall this just tensor products with the function output. Upon measurement on those function qubits as $f(r)$, we get the desired output in the 1st register (i.e. first part of the output):

$$\frac{1}{\sqrt{2^n}}\sum_x |x\rangle|f(x)\rangle \mapsto \frac{1}{\sqrt{2}}|r\rangle + \frac{1}{\sqrt{2}}|r \oplus s\rangle$$

This because upon measuring $f(r)$, the value in the 1st register is $|r\rangle$ or $|r \oplus s\rangle$ with equal probability.

Thus, the circuit looks as follows:



Now, let's make sure the claim in step 2 is correct. Applying the Hadamard transform yields the following:

$$\begin{aligned} H^{\otimes n}\left(\frac{1}{\sqrt{2}}|r\rangle + \frac{1}{\sqrt{2}}|r \oplus s\rangle\right) &= \frac{1}{\sqrt{2^{n+1}}}\sum_a \left((-1)^{a\cdot r} + (-1)^{a\cdot(r\oplus s)}\right)|a\rangle \\ &= \frac{1}{\sqrt{2^{n+1}}}\sum_a (-1)^{a\cdot r}(1 + (-1)^{a\cdot s})|a\rangle \end{aligned}$$

However, note that:

$$(1 + (-1)^{a\cdot s}) = \begin{cases} 0 & \text{if } a \cdot s = 1 \\ 2 & \text{if } a \cdot s = 0 \end{cases}$$

So we finally get the state as:

$$\frac{1}{\sqrt{2^{n-1}}} \sum_{a:a\cdot s=0} (-1)^{a\cdot r} |a\rangle$$

Measuring ignores the phase, so we thus have a uniform distribution over all the 2^{n-1} vectors orthogonal to s.

Finally, note that the measurement made the math a lot easier, but by the principle of deferred measurement the measurement on the second n qubits is not strictly necessary.

7.2 Quantum Fourier Transform

Recall the roots of unity over $\mathbb{C}$.

Note 7.1 (Roots of Unity)
An Mth root of unity is a complex number z such that $z^M = 1$. The primitive Mth root of unity ω is

$$\omega = e^{2\pi i/M} = \cos\frac{2\pi}{M} + i\sin\frac{2\pi}{M}$$

Furthermore, any Mth root of unity can be written as ω^k for $0 \le k < M$.

The Fourier transform is just the act of applying a polynomial on roots of unity. Suppose you have a polynomial $\alpha(x) = \sum_{j=0}^{M-1} \alpha_j x^j$. Then we can write $\beta_k = \alpha(\omega_k)$, which can be expanded and written as multiplication by a special Vandemonde matrix, called the **Discrete Fourier Transform (DFT)**.

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{M-1} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \omega & \dots & \omega^{M-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \omega^{M-1} & \dots & \omega^{(M-1)(M-1)} \end{pmatrix}}_{\text{DFT Matrix}}$$

We add a normalizing factor because we want to use quantum bases, so we'll work with a slightly different definition:

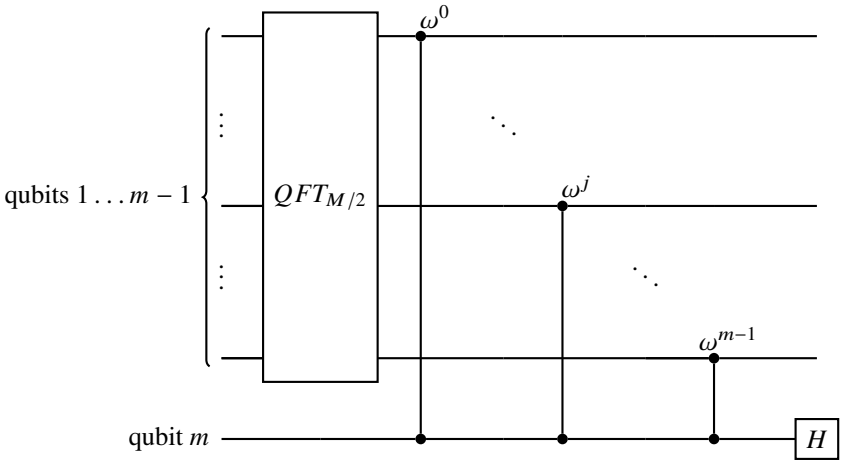
$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{M-1} \end{pmatrix} = \frac{1}{\sqrt{M}} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \omega & \dots & \omega^{M-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \omega^{M-1} & \dots & \omega^{(M-1)(M-1)} \end{pmatrix}$$

The naive algorithm to do this matrix-vector multiplication is $O(M^2)$, but there is a divide-and-conquer algorithm called the **Fast Fourier Transform (FFT)** that can do this much faster—in $O(M \log M)$ time. The **Quantum Fourier Transform (QFT)** can do this in $\tilde{O}(\log M)$, where $\tilde{O}$ hides poly-log factors. We will show a simple way to achieve $O(\log^2 M)$. However, there is a big caveat: we know the answer takes M time to even write down. The reason QFT can go faster is that since the QFT gives you a quantum state, you only get a single index j upon any actual measurement; the FFT gives you the entire answer!

Now we discuss the implementation of the QFT, which will be surprisingly similar to the FFT. Without loss of generality, assume M is a power of two and $m = \log_2 M$. By the matrix multiplication above, it's clear that:

$$|k\rangle \overset{QFT}{\mapsto} \frac{1}{\sqrt{M}} \sum_{j=0}^{M-1} \omega^{jk} |j\rangle =: |\chi_k\rangle$$

Suppose inductively that we knew how to apply the $QFT_{M/2}$. Then to apply QFT_M we can implement it as:



Note that unrolling the recursion, and calling QFT_0 the identity, it's clear that the runtime is:

$$1 + 2 + \cdots + (m-1) + m = O(m^2) = O(\log^2 M)$$

Now to prove correctness, consider the FT matrix. We reorder the columns so the first $M/2$ columns are the even-indexed columns (0-indexing) and the rest are odd-indexed columns. Then

$$FT_M = \begin{pmatrix} FT_{M/2} & \omega^j FT_{M/2} \\ FT_{M/2} & -\omega^j FT_{M/2} \end{pmatrix}$$

where $\alpha^j A$ means to multiply the jth row of A by α^j. To show this, let's consider the four quadrants. Call j the row, ℓ the column, and k some integer satisfying $0 \le k < M/2$. We will only do two cases for simplicity, the rest are similar. If you're in the top left, e.g. if $0 \le j < M/2$ and $\ell = 2k$, then the entry is $\omega_M^{2jk} = \omega_{M/2}^{jk}$, i.e. the correct entry in the top left. If you're in the bottom right, e.g. if $M/2 \le j < M$ and $\ell = 2k+1$. Let $j' = j - M/2$, then the entry is

$$\omega^{(2k+1)(j'+M/2)} = \omega^{2kj'+kM+j'+M/2} = \omega^{kM}\omega_{M/2}^{kj'}\omega^{M/2+j'} = -\omega^{j'}\omega_{M/2}^{kj'}$$

This proves the claim, allowing us to use divide-and-conquer to compute the full FT efficiently.

8 Lecture 8

8.1 Factoring

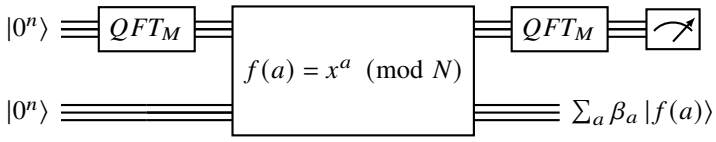
We discuss the classic factoring problem. Given a number N, we wish to find its prime factorization

$$N = p_1^{e_1} p_2^{e_2} \dots p_k^{e_k}$$

The simplest algorithm is to try dividing N by every number up to its square root (its biggest factor). This will take $O(\sqrt{N})$ time. But if you call n the number of bits to write N, e.g. about $\log N$, then $\sqrt{N} = 2^{n/2}$. So this algorithm is still exponential in n. If we are in a cryptographic setting, where n is a 1024-bit or 2048-bit number, $\sqrt{N}$ is gigantic. The best known classical algorithm is the Quadratic Number Sieve, which runs in $O(\exp(\sqrt[3]{n}))$. We would prefer a $\text{poly}(n)$ algorithm (or even better to disprove its existence to protect cryptography). It turns out we can do better with a quantum computer.

If we could just factor a composite N into two non-unit factors, $N = N_1 \cdot N_2$, then we could easily factor it by repeated calling this algorithm (the divide and conquer perhaps gaining an extra poly-logarithmic factor). Then, the hardest case is really when $N = P \cdot Q$ for two primes of roughly equal size.

Here is a circuit we claim splits N into such factors. Fix $M > N$ and pick x at random satisfying $0 < x < N$.



Then from the measurement, we do a little bit of classical postprocessing to yield our two factors.

Let's work with an example to motivate how this circuit works. We have $N = 15, M = 16$, then for chosen $x = 2$:

$$x^0 \equiv x^4 \equiv x^8 \equiv x^{12} \equiv 1 \pmod{15}$$

Furthermore

$$x^1 \equiv x^5 \equiv x^9 \equiv x^{13} \equiv 2 \pmod{15}$$

If $\gcd(x, N) = 1$, then there exists a smallest $r > 0$ such that $x^r \equiv 1 \pmod N$. We call r the order of x. If we picked an x that isn't relatively prime, we could use the Euclidean algorithm to find a common factor of x and N and thus a factor of N.

Suppose we knew the order of x mod 15 was 4 beforehand. Since r is even, let's try halving it (if it wasn't, we'd need a different x). So, define $y \equiv x^{r/2} \pmod{15}$. This means that it's a second root of unity, e.g. $y^2 \equiv 1 \pmod{15}$. In the special case when N is a product of two primes, there are four roots of unity. In this case we have $y = 2^2 = 4$, which is not trivially plus or minus 1 (if it was, we'd need a different x). Now, this means

$$y^2 - 1 \equiv 0 \pmod{15} \implies 15 \mid (y+1)(y-1)$$

But we know that 15 cannot divide either thing individually, since $y \not\equiv \pm 1 \pmod{15}$. Thus, now we can just compute $\gcd(15, y+1)$ and $\gcd(15, y-1)$; this gives us our two factors. In our example this is $\gcd(15, 5) = 5$ and $\gcd(15, 3) = 3$. We claim that our circuit above will find the order of x.

8.2 Period Finding

How do we extract the order of x (e.g. the period)? Now let's think about what we could get out of the circuit right before the second QFT based on different measurements of $f(a)$ (which we can consider by the principle of deferred measurement).

$f(a)$	State of first register
0	Not possible
1	$\|0\rangle + \|4\rangle + \|8\rangle + \|12\rangle$
2	$\|1\rangle + \|5\rangle + \|9\rangle + \|13\rangle$
3	Not possible
4	$\|2\rangle + \|6\rangle + \|10\rangle + \|14\rangle$

Suppose r divides M for now. Then, the superposition we get (up to normalization) is

$$|\psi\rangle = \sum_{j=0}^{\frac{M}{r}-1} \sqrt{\frac{r}{M}} |jr\rangle \overset{QFT_M}{\mapsto} \sum_{\ell=0}^{M-1} \sum_{j=0}^{\frac{M}{r}-1} \frac{\sqrt{r}}{M} \omega^{jr\ell} |\ell\rangle$$

For $\ell = \frac{kM}{r}$, we get a constructive interference of M/r terms, each with amplitude $\frac{\sqrt{r}}{M}$, which gives overall amplitude $\frac{1}{\sqrt{r}}$. But now notice that there are r such terms, which means the superposition is normalized by these; the other amplitudes must be zero. This means that

$$|\psi\rangle \overset{QFT_M}{\mapsto} \frac{1}{\sqrt{r}} \sum_{\ell=0}^{r-1} \left| \frac{M}{r} \ell \right\rangle$$

Now, if we continually measure, we can get the period with high probability, simply by finding $\frac{M}{r}$ via the gcd of the measurements. In other words, we can find the period by solving

$$T = \gcd(\ell_1, \ell_2, \dots) = \gcd\left(\frac{k_1 M}{R}, \frac{k_2 M}{R}, \dots \right) = \frac{M}{r}.$$

One can view this as the time-frequency uncertainty principle: increasing the period r in the original domain decreases the period T in the new domain.

Note that everything we've discussed so far holds only when r divides M. In the general case, r does not divide M. But with constant probability, we see ℓ that satisfies $|\ell r \mod M| \leq r/2$, i.e.,

$$\left| \frac{\ell}{M} - \frac{k}{r} \right| \leq \frac{1}{2M}$$

If we choose $M \approx N^2$, then we can figure out r by tightening this bound. The continued fractions algorithm does the job.

9 Lecture 9

9.1 Grover's Algorithm

We investigate the unstructured search problem. Suppose we have a boolean function $f : \{0,1\}^n \to \{0,1\}$, which is guaranteed to not be the 0 function. We wish to find an a such that $f(a) = 1$. Call $N = 2^n$. One can think of a Boolean function as a table, mapping numbers from 0 to $N-1$. Deterministically you must brute force all entries in the worst case, so it takes N; with a randomized algorithm, on average you must look through half of them $N/2$, which is still exponential in n. This problem is hard (in *NP-Complete*), as it's easy to check given an a that $f(a) = 1$, but it's hard to find that a. We discuss the hardest version of the problem, where f is only 1 at a single point.

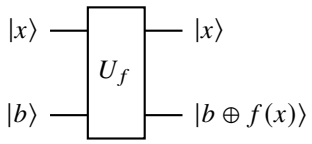
In the quantum world, we can prepare the following superposition:

$$\frac{1}{\sqrt{N}} \sum_x |x\rangle |f(x)\rangle$$

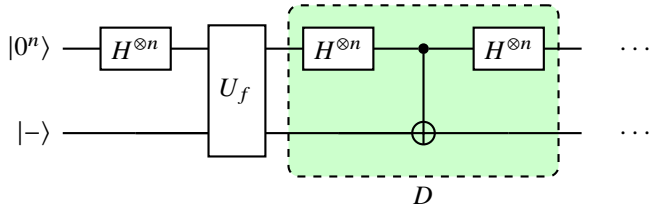
However, when we measure, we only get 1 with probability $\frac{1}{N}$, and we're no better than classical. In fact, the best we can do is still exponential.

Theorem 9.1 (BBBV Lowerbound)
Any quantum algorithm for the unstructured search problem must take $\gtrsim \sqrt{N}$ time.

Grover's algorithm saturates this lower bound of $O(\sqrt{N})$. Recall that one thing we can do with quantum computation is create a phase state: $\frac{1}{2^{n/2}} \sum_x (-1)^{f(x)} |x\rangle$. Consider the bit flip operator X; its eigenvectors are $|\pm\rangle$ with eigenvalues ± 1. Consider the following circuit which computes $f(x)$ and applies a CNOT to $|f(x)\rangle$ and $|b\rangle$:



Now imagine if $|b\rangle$ is $|-\rangle$. Then, our output on the right is $(-1)^{f(x)} |-\rangle$ by linearity. This creates the phase state we want (we can throw away $|-\rangle$ since it's in a tensor product). This leads to the following circuit, where D is called the diffusion operator.



What does it mean to control on a bunch of bits? In this case, this means that if all of the bits are 0, then leave it alone, otherwise, we flip. Everything after the first Hadamard is considered "one iteration" of Grover's algorithm. We run this for $O(\sqrt{N})$ iterations, and we measure at the end. We claim that a pops out with constant probability (with respect to N).

Note 9.1
The diffusion operator D can be expressed as follows:

$$D = \frac{2}{N}J - I = \begin{pmatrix} \frac{2}{N} - 1 & \frac{2}{N} & \cdots & \frac{2}{N} \\ \frac{2}{N} & \frac{2}{N} - 1 & \cdots & \frac{2}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{2}{N} & \frac{2}{N} & \cdots & \frac{2}{N} - 1 \end{pmatrix}$$

where J is the all 1's matrix.

Let's look at what the diffusion operator actually does. Consider

$$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{N-1} \end{pmatrix} = D\alpha = D \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{N-1} \end{pmatrix},$$

where each entry of β is $\beta_i = -\alpha_i + \frac{2}{N}\sum_j \alpha_j$. Then, noting that the mean is $\mu = \frac{1}{N}\sum_j \alpha_j$, we can express each of these entries as

$$\beta_i = 2\mu - \alpha_i = \mu - (\alpha_i - \mu).$$

Now, we see that applying D inverts every element about the mean. Let's use this property of D to analyze the correctness of Grover's algorithm.

9.1.1 Analysis of Grover's Algorithm

At step 0, all the amplitudes are evenly distributed as $\frac{1}{\sqrt{N}}$. Applying U_f, we end up with the phase state, which is -1 at a and 0 everywhere else. This causes the mean of the amplitudes to decrease ever so slightly, so that after applying D, the rest of the amplitudes shift a small amount, and a's amplitude goes up to nearly $\frac{3}{\sqrt{N}}$. After another iteration, it goes up to $\frac{5}{\sqrt{N}}$, and it keeps on increasing by about $\frac{2}{\sqrt{N}}$ as long all the amplitudes are close to $\frac{1}{\sqrt{N}}$. In fact, we get an increase of at least $\frac{1}{\sqrt{N}}$ as long as the other amplitudes are above $\frac{1}{2\sqrt{N}}$.

Thus, after $O(\sqrt{N})$ steps, the amplitude of a takes up a constant fraction of the norm (in particular, the rest take up probability mass $\left(\frac{1}{2\sqrt{N}}\right)^2 \cdot (N-1) \leq \frac{1}{4N} \cdot N = \frac{1}{4}$). This means that we can measure a with constant probability (at least $\frac{3}{4}$), as desired.

9.1.2 Geometric Interpretation of Grover's Algorithm

Here is a geometric way to view what Grover's algorithm does. Consider the two-dimensional subspace spanned by $|a\rangle$ and $|u\rangle = \sum_x \frac{1}{\sqrt{N}}|x\rangle$. Note that these two vectors are not orthogonal, so we introduce $|e\rangle$, which is a vector orthogonal to $|a\rangle$ in this plane. Now, we can interpret this plane to have basis $\{|e\rangle, |a\rangle\}$ (i.e. the x, y axes), and $|u\rangle$ is a vector lying within this 2D coordinate system. Then, geometrically, this is what Grover's does:

1. Start in the state $|u\rangle$.
2. Reflect about $|e\rangle$ (i.e. applying U_f).
3. Reflect about $|u\rangle$ (i.e. applying D).
4. Repeat steps 2-3 for each subsequent iteration.

One can think of steps 2-3 as walking $|u\rangle$ towards $|a\rangle$. Initially, the angle between $|u\rangle$ and $|e\rangle$ is $\theta = \sin^{-1}\frac{1}{\sqrt{N}} \approx \frac{1}{\sqrt{N}}$. After one iteration, the new angle becomes $\theta + 2\theta = 3\theta$, and each subsequent iteration increases the angle by 2θ. Thus, in order to get to $\theta' = \frac{\pi}{2}$ (which corresponds to $|a\rangle$), it takes approximately

$$\frac{\pi/2}{2\theta} \approx \frac{\pi/2}{2/\sqrt{N}} = \frac{\pi}{4}\sqrt{N} = O(\sqrt{N})$$

iterations.

9.2 Vaidman's Bomb

Suppose there is a quantum-superposition of bomb. If we look inside and there is a bomb, then it will blow up, but if it's safe, then we want to open it. Consider sending a photon through it with state $|\psi\rangle = \cos k\theta\,|0\rangle + \sin k\theta\,|1\rangle$, where the $|0\rangle$ state is a photon that passes through without looking at the bomb and $|1\rangle$ state is a photon that passes through while checking for the bomb. You send this state through the box. If we measure the photon on the outside and there is a bomb, the probability of an explosion is $\sin^2\theta$. If there isn't a bomb, then for sure we will not explode. We repeat this N times by rotating by θ such that $N\theta = \frac{\pi}{2}$. Then the Nth measurement is definite and tells you there is no bomb. The probability that we explode every time is at most θ^2; taking a union bound makes it at most $N\theta^2 = O(1/N)$.

This means intuitively, we make $1/N$ progress towards the probability 1 every time.

Part II
Harnessing Quantum Systems

10 Lecture 10

The previous lectures assumed that quantum systems were under our total control, which is completely unrealistic. For the next 50 years or so, there will likely be no meaningful systems that are completely within our control. Thus, we turn our attention to figuring out how quantum systems act on their own, describing quantum noise and quantum error correction, which has far-reaching consequences beyond just computing. In our next chapter, we will discuss real-life experimental implementations of quantum computers in labs.

10.1 Schrodinger's Equation

As we claimed earlier, the (time) evolution of a quantum system is unitary.

$$|\psi(t_2)\rangle = U(t_2, t_1)\,|\psi(t_1)\rangle$$

Furthermore, we should be able to compose these operators:

$$|\psi(t_3)\rangle = U(t_3, t_2)U(t_2, t_1)\,|\psi(t_1)\rangle \implies U(t_3, t_1) = U(t_3, t_2)U(t_2, t_1)$$

Namely, we must have consistency of the quantum circuit. Thus, to know the evolution of our system, it is sufficient to know how to take small steps of ϵ, i.e. we only need $U(t+\epsilon, t)$. Intuitively, we should expect that the dynamics are small, i.e. Taylor expandable as:

$$U(t+\epsilon, t) = I - i\epsilon H(t) + O\left(\epsilon^2\right)$$

But we know this matrix is unitary, so:

$$I = U^\dagger U = \left(I + i\epsilon H^\dagger + O\left(\epsilon^2\right)\right)\left(I - i\epsilon H + O\left(\epsilon^2\right)\right) = I + i\epsilon(H^\dagger - H) + O\left(\epsilon^2\right)$$

For the left and right hand sides to agree, we must have $i\epsilon\left(H - H^\dagger\right) = 0$, or $H = H^\dagger$. This means H is a **Hermitian matrix**, which we can think of as a generalization of a symmetric matrix to the complex case. This means we can write our evolution as approximately:

$$\begin{aligned} |\psi(t+\epsilon)\rangle &= (I + i\epsilon H(t))\,|\psi(t)\rangle \\ \frac{|\psi(t+\epsilon)\rangle - |\psi(t)\rangle}{\epsilon} &= -iH(t)\,|\psi(t)\rangle \\ i\frac{\partial}{\partial t}\,|\psi(t)\rangle &= H(t)\,|\psi(t)\rangle \end{aligned}$$

where we took a limit as $\epsilon \to 0$. This equation is referred to as **Schrodinger's Equation**, with unit $\hbar = 1$. We are often interested in the special case where H is a constant, i.e. the dynamics are time-homogenous with $U(t_2, t_1) = U(t_2 - t_1, 0)$. This operator H has a special name, the **Hamiltonian** of the system. In fact, it is the measurement operator associated to the energy of the system.

Definition 10.1
A measurement operator is a Hermitian operator A for an observable O with (orthonormal) eigenstates $|\psi_n\rangle$ and eigenvalues A_n, where the former are states of definite O value and A_n are the values of O you can observe.

We are guaranteed the eigenstates of the Hermitian operator are orthonormal by the Spectral theorem, so this works with our idea of measurement before. For the Hamiltonian, this means that we can refer to the eigenvectors $|\psi_n\rangle$'s as the states with definite energy, and the E_n's as the energies of the corresponding states. They are related them via

$$H\,|\psi_n\rangle = E_n\,|\psi_n\rangle\,,$$

which is often called the **time-independent Schrodinger equation**. With this substitution, we obtain

$$i\frac{\partial}{\partial t}|\psi(t)\rangle = E_n|\psi(t)\rangle \implies |\psi(t)\rangle = e^{-iE_nt}|\psi(0)\rangle = e^{-iE_nt}|\psi_n\rangle$$

So essentially, Schrodinger's Equation tells us that the behavior for a given quantum state $|\psi(t)\rangle$ is characterized by its component states $|\psi_n\rangle$ precessing at a rate proportional to its energy E_n.

Example 10.1
Given initial state $|\psi(0)\rangle = |0\rangle$ and operator $H = X = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$, find the general state $|\psi(t)\rangle$.

Solution: Since the eigenvector-value pairs of H are $(|+\rangle, 1)$ and $(|-\rangle, -1)$, we express $|\psi(0)\rangle$ in terms of the eigenvectors of H as follows:

$$|\psi(0)\rangle = |0\rangle = \sum_n \alpha_n|\psi_n\rangle = \frac{1}{\sqrt{2}}|+\rangle + \frac{1}{\sqrt{2}}|-\rangle$$

Now, we can compute $|\psi(t)\rangle$ as such:

$$\begin{aligned}|\psi(t)\rangle &= e^{-iHt}|\psi(0)\rangle\\ &= \sum_n \alpha_n e^{-iE_nt}|\psi_n\rangle\\ &= \frac{1}{\sqrt{2}}e^{-i\cdot 1\cdot t}|+\rangle + \frac{1}{\sqrt{2}}e^{-i\cdot(-1)\cdot t}|-\rangle\\ &= \frac{1}{\sqrt{2}}e^{-it}|+\rangle + \frac{1}{\sqrt{2}}e^{it}|-\rangle\end{aligned}$$

where we choose units such that $\hbar = 1$.

10.2 Bloch Sphere

Recall we tried to visualize $\alpha|0\rangle + \beta|1\rangle$ as living in a 2d Hilbert space. However, in general, the amplitudes are complex, so such a picture is incomplete.

For the special case of a two-state system (e.g. a qubit), there exists a different visualization technique. Note that even though there are 2 complex numbers and thus 4 real numbers of information, scaling by phase and magnitude does not change the state, i.e. $|\psi\rangle \equiv \lambda|\psi\rangle$ for any $\lambda \in \mathbb{C}$. Thus, there are only $4 - 2 = 2$ real degrees of freedom. We normalize the state, taking the norm to be 1, and use the following form:

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

Now we can remove the phase of $e^{i\chi}$ from each, so we get the state as

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

where $\phi \in [0, 2\pi]$ and $\theta \in [0, \pi]$ (one period of all of these functions). If $\theta = 0$, then clearly ϕ is irrelevant because the second term is 0. Similarly, if $\theta = \pi$, then clearly there is only a $|1\rangle$ term, which means that ϕ only creates a global phase, which doesn't change the state.

Hence, we call this the **Bloch Sphere** representation of the state $|\psi\rangle$. This is because it naturally maps a state to points on a sphere, with θ as the polar angle and ϕ as the azimuthal angle in a standard spherical coordinate setup. The special cases above are the north ($|0\rangle$) and south pole ($|1\rangle$) separately, where at a pole your azimuthal angle doesn't matter.

Note that the poles we marked aren't special, and we can generalize the idea of "poles" to **antipodal points**:

Definition 10.2 (Antipodal)
Two points on the Bloch sphere are defined to be antipodal if the line segment connecting them crosses the center of the sphere.

Note 10.1
Given a state $|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle$, the state that is antipodal to it is

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

Note 10.2
Orthogonal states always occur at antipodal (opposite) points on the Bloch sphere.

One can also observe that the states $|\pm\rangle$ and $|\pm i\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle)$ all live on the equator with $\theta = \frac{\pi}{2}$, where $|\pm\rangle$ points in the direction of $\pm x$ and $|\pm i\rangle$ points in the direction of $\pm y$.

10.3 Electron Spins

One can think of electron spins as a two state system with basis $|\uparrow\rangle = |0\rangle$ and $|\downarrow\rangle = |1\rangle$ which correspond (roughly) to the direction an electron can spin (angular momentum in the $\pm z$ axis). You can also pick any direction $\mathbf{n}$ and find the component of the angular momentum in that direction.

Classically, the spin will just be $\mathbf{n}\cdot\mathbf{z}$ and it will be some definite value. However, in the world of quantum mechanics, by measuring in this direction, we can only get two possibilities: the spin being either parallel or antiparallel with $\mathbf{n}$. One can use the Bloch sphere to understand this; associate $|\mathbf{n}\rangle$ with its representation on the Bloch sphere. The state $|\mathbf{n}\rangle$ has definite angular momentum $L_{\mathbf{n}} = +\frac{1}{2}$ in the $\mathbf{n}$-direction, so it and its antipodal point $|-\mathbf{n}\rangle$ also form a basis for two state systems!

We can think of the operator for the measurement of momentum in the z direction (i.e. the "observable") as follows:

$$L_{\mathbf{z}} = \frac{1}{2}|0\rangle\langle 0| - \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} = \frac{1}{2}Z$$

Similarly, for measuring the x momentum we have

$$L_{\mathbf{x}} = \frac{1}{2}|+\rangle\langle +| - \frac{1}{2}|-\rangle\langle -| = \frac{1}{2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} = \frac{1}{2}X$$

We know from physics that a magnetic field assigns a vector $\mathbf{B}$ to every point in space. Then, we can write the Hamiltonian (total energy) as:

$$H = \frac{e}{m_e}\mathbf{L}\cdot\mathbf{B}$$

where e and m_e are the charge and mass of an electron, respectively. For simplicity, we will say the magnetic field is constant and only in the x direction: $\mathbf{B} = (B, 0, 0)$, which means

$$H = \frac{e}{m_e}BL_x = \frac{eB}{2m_e}X.$$

Now that we know the Hamiltonian, we can figure out how such a system evolves in time. For our initial condition, we set $|\psi(0)\rangle = |\uparrow\rangle = |0\rangle$, and write the general state as $|\psi(t)\rangle = \alpha(t)|0\rangle + \beta(t)|1\rangle$. Rewriting Schrodinger's equation we

get:

$$i\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = H|\psi(t)\rangle$$
$$i\dot{\alpha}(t)|0\rangle + i\dot{\beta}(t)|1\rangle = \frac{eB}{2m_e}(\alpha(t)X|0\rangle + \beta(t)X|1\rangle)$$
$$i\dot{\alpha}(t)|0\rangle + i\dot{\beta}(t)|1\rangle = \frac{eB}{2m_e}(\beta(t)|0\rangle + \alpha(t)|1\rangle)$$

This gives us a differential equation for each component:

$$\begin{cases}\dot{\alpha} = \frac{eB}{2m_e i}\beta \\ \dot{\beta} = \frac{eB}{2m_e i}\alpha\end{cases}$$

Taking the derivative with respect to t in the first equation, we get that

$$\ddot{\alpha} = \frac{eB}{2m_e i}\dot{\beta} = -\left(\frac{eB}{2m_e}\right)^2\alpha$$

This is a well-known ODE with solutions as sin and cos. Note that α must be a cosine because it is initially 1, so working it out gives us:

$$\alpha(t) = \cos\left(\frac{eB}{2m_e}t\right) \implies \beta(t) = \frac{2m_e}{eB}i\dot{\alpha} = -i\sin\left(\frac{eB}{2m_e}t\right)$$

We can interpret this behavior as rotating around the bloch sphere, where the relative phase ϕ stays fixed at π (or flips as we go back around) and θ moves at the angular velocity $\omega = \frac{eB}{2m_e}$, with period $\frac{\pi m_e}{eB}$. The electron rotates down to $|+i\rangle$, then $|1\rangle$, and finally $|-i\rangle$. This is a rotation about the x-axis, where we put the magnetic field! This behavior is known as **spin precession**.

10.3.1 Using Spin to Enable Quantum Computing

How does this idea enable quantum computing? We know that we have to somehow change the Hamiltonian throughout time. To do this, we implement quantum gates by switching the magnetic field on and off in certain directions to create desired electron motion. For example, to implement a NOT gate, we leave magnetic field on for some multiple of the period (plus one half phase) to make the spin precess until it hits the opposite pole, then switch off the magnetic field. This way, we're able to manipulate the behavior of quantum states.

11 Lecture 11

11.1 Describing Noisy Quantum States

Recall that in a (noisy) classical system, we define a probability distribution over N states $\{p_i\}_{i=1}^N$ where each p_i tells you the probability of being in state i. Similarly, in a quantum system, we can define probabilities p_i of being in a state $|\psi_i\rangle$. Suppose we measure in the $|m\rangle$ basis, so the probability of measuring m is

$$\mathbb{P}[m] = \sum_i p_i \mathbb{P}[m \mid |\psi_i\rangle] = \sum_i p_i |\langle m|\psi_i\rangle|^2$$

But, we can rewrite this quadratic term as a trace:

$$|\langle m|\psi_i\rangle|^2 = \langle\psi_i|m\rangle \langle m|\psi_i\rangle = \mathrm{Tr}\left(|\psi_i\rangle\langle\psi_i| \, |m\rangle\langle m|\right)$$

Using the fact that trace is linear, we can see:

$$\mathbb{P}[m] = \mathrm{Tr}\left(\left(\sum_i p_i |\psi_i\rangle\langle\psi_i|\right)|m\rangle\langle m|\right) = \mathrm{Tr}(\rho \, |m\rangle\langle m|)$$

where we define the **density operator** as:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

Note that ρ consists of d^2 matrix elements, where d is the dimension of the Hilbert space. However, the space of ensembles is infinite-dimensional, because there are an infinite amount of states and we can define any probability distribution we'd like on them. The reason for this is because different ensembles can have the same ρ operator.

Example 11.1
Take the probability distribution $p(|0\rangle) = 0.5$, $p(|1\rangle) = 0.5$, which gives us density

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

But now take the distribution $p(|+\rangle) = 0.5$, $p(|-\rangle) = 0.5$, which gives us the same density

$$\rho = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| = \frac{1}{2}I$$

Seeing this, it is natural to say that a density matrix ρ defines a noisy quantum state; if $p_i \neq 0$ for more than one i, then this is called a **mixed state**. If you have two systems with the same density matrix, they are indistinguishable. A state $|\psi\rangle \in \mathcal{H}$ is called a **pure state**, with density matrix $\rho = |\psi\rangle\langle\psi|$. One may note that this is a **projector**, i.e. an orthogonal projection matrix which takes a state and projects it onto its component in the ψ direction. Formally:

Definition 11.1 (Projector)
A projector P is a Hermitian matrix with only two eigenvalues, 0 and 1, i.e.

$$P = P^\dagger = P^2$$

The nice thing about this matrix is that we don't need to worry about the global phase indistinguishability, as that will cancel in the projector expression.

11.2 Properties of the Density Operator

We have that

$$\rho^\dagger = \sum_i p_i (|\psi_i\rangle\langle\psi_i|)^\dagger = \sum_i p_i |\psi_i\rangle\langle\psi_i| = \rho$$

Thus, ρ is Hermitian. Furthermore, it is actually positive semi-definite:

$$\langle\psi_i|\,\rho\,|\psi_i\rangle = \sum_i p_i |\langle\psi_i|\psi_i\rangle|^2 \geq 0$$

Thus, its eigenvalues are all nonnegative. Finally, we have that

$$\mathrm{Tr}(\rho) = \sum_i p_i = 1$$

because the p_i's form a probability distribution.

Theorem 11.1
Any trace-1 PSD matrix ρ is the density matrix of infinitely many ensembles $\{(p_i, |\psi_i\rangle)\}$.

Proof
By the spectral theorem, we can write such a matrix as

$$\rho = \sum_j \lambda_j \left|\psi_j\right\rangle\!\left\langle\psi_j\right|$$

Since ρ is PSD, then $\lambda_j \geq 0$ and since it has trace 1, $\sum_i \lambda_i = 1$. Thus, this is the density operator for the ensemble $\{(\lambda_j, |\psi_j\rangle)\}$.
By uniqueness of eigendecomposition, this is in fact the unique ensemble in which all the states are orthonormal.

11.3 QM Postulates in terms of Density Matrices

Now, let's visit the postulates of Quantum Mechanics in terms of density matrices.

Theorem 11.2

1. **Superposition**: For pure states, the original superposition principle states that states live in a Hilbert space $|\psi\rangle \in \mathcal{H}$. For density operators, we'd say that a state is described as a PSD trace-one matrix ρ acting on $\mathcal{H}$.

2. **Measurement**: For pure states, recall that we measure in an orthonormal basis $\{|m\rangle\}$ such that $\mathbb{P}[m] = |\langle\psi|m\rangle|^2$ and $|\psi\rangle$ collapses to $|m\rangle$. But to deal with more general measurements, e.g. measuring on a certain qubit, we use a **projection-valued measure (PVM)**. We say we have a set of projectors P_m such that $\sum P_m = I$. For a measurement on a one-qubit state, we use $P_m = |m\rangle\!\langle m|$; and for a measurement on the first qubit but not the second, we use $P_m = |m\rangle\!\langle m| \otimes I$, wherein measurement yields

$$\mathbb{P}[m] = \langle\psi|P_m|\psi\rangle\,, \text{ and we update } |\psi\rangle \to \frac{P_m|\psi\rangle}{\sqrt{\langle\psi|P_m|\psi\rangle}}$$

   Now for density operators, measurement turns into

$$\mathbb{P}[m] = \sum_i p_i \langle\psi_i|P_m|\psi_i\rangle = \mathrm{Tr}(\rho P_m)$$

   The update rule is a bit harder. Measuring m gives information about which $|\psi_i\rangle$ we were in, so we have

$$\rho \to \sum_i \mathbb{P}[|\psi_i\rangle \mid m] \frac{P_m|\psi_i\rangle\langle\psi_i|P_m}{\langle\psi_i|P_m|\psi_i\rangle}$$

   To find the conditional probability that we were in state ψ_i given a measurement of m, we compute

$$\mathbb{P}[|\psi_i\rangle \mid m] = \frac{p_i \mathbb{P}[m \mid |\psi_i\rangle]}{\mathbb{P}[m]} = \frac{p_i \langle\psi_i|P_m|\psi_i\rangle}{\mathrm{Tr}(\rho P_m)}$$

This means the final update is:

$$\begin{aligned}\rho &\to \sum_i \left[\frac{p_i \langle\psi_i|P_m|\psi_i\rangle}{\mathrm{Tr}(\rho P_m)} \cdot \frac{P_m |\psi_i\rangle \langle\psi_i| P_m}{\langle\psi_i| P_m |\psi_i\rangle}\right] \\ &= \frac{P_m \left(\sum_i p_i |\psi_i\rangle\langle\psi_i|\right) P_m}{\mathrm{Tr}(\rho P_m)} \\ &= \frac{P_m \rho P_m}{\mathrm{Tr}(\rho P_m)}\end{aligned}$$

One can think about this as follows: project the operator ρ into the subspace spanned by P_m, then "normalize" the trace by using the denominator.

3. **Unitary Evolution**: For pure states, we evolve a state with a unitary by applying $|\psi\rangle \mapsto U|\psi\rangle$. For mixed states, we have:
$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \mapsto \sum_i p_i U |\psi_i\rangle\langle\psi_i| U^\dagger = U\rho U^\dagger$$

11.4 Geometric Interpretation of Density Matrices

With that formal stuff out of the way, let's think about a two state system. The spectral decomposition of such an operator must have only two eigenvectors and with the trace-one condition we have:

$$\rho = \lambda |\psi\rangle\langle\psi| + (1-\lambda)(I - |\psi\rangle\langle\psi|)$$

Without loss of generality, we can take $\frac{1}{2} \le \lambda \le 1$ because we can just pick the biggest we can parametrize $\lambda = \frac{1}{2}(1+r)$ where $0 \le r \le 1$ and we can think of mixed states geometrically. Recall our parametrization of the Bloch sphere last lecture in terms of θ and ϕ:

1. If $r = 1$, we get a pure state $|\psi\rangle\langle\psi|$.
2. If $r = 0$, then $\lambda = 1/2$ and the expression at the top is $\rho = \frac{1}{2}I$ regardless of the $|\psi\rangle$ we choose.

Other mixed states lie in the inside of the sphere; the closer you are to the center, the noisier your state. Hence, any $\rho \propto I$ (in the center of the sphere) is called a "maximally mixed" state.

11.5 Reduced Density Matrices

Suppose you have two coupled systems $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. One can think of $\mathcal{H}_A$ as a quantum computer or some other quantum system we wish to control; $\mathcal{H}_B$ is any other noise/system that $\mathcal{H}_A$ can potentially interact with. Given some $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ we wish to only study its state on $\mathcal{H}_A$. But maybe we can write $|\psi\rangle = |\psi_A\rangle |\psi_B\rangle$—then we'd be very lucky.

What do measurements on A look like? They look like $P_m^A \otimes I^B$. The probability of observing m is:

$$\mathbb{P}[m] = \langle\psi|P_m^A \otimes I^B|\psi\rangle = \mathrm{Tr}\left[\left(P_m^A \otimes I^B\right)|\psi\rangle\langle\psi|\right]$$

Pick some orthonormal bases $|i\rangle_A, |j\rangle_B$ for the two spaces. We then have:

$$\begin{aligned}\mathbb{P}[m] &= \sum_{i,j} \langle i|_A \langle j|_B P_m^A \otimes I^B |\psi\rangle\langle\psi| |i\rangle_A |j\rangle_B \\ &= \sum_i \langle i|_A P_m^A \left(\sum_j \langle j|_B |\psi\rangle \langle\psi|j\rangle_B\right) |i\rangle_A \\ &= \sum_i \langle i|_A P_m^A \,\mathrm{Tr}_B(|\psi\rangle\langle\psi|)\, |i\rangle_A\end{aligned}$$

where we define

$$\mathrm{Tr}_B(X) = \sum_j \langle j|_B \, X \, |j\rangle_B$$

Now, define the **reduced density matrix** $\rho_A = \mathrm{Tr}_B(|\psi\rangle\langle\psi|)$ (in general $\rho_A = \mathrm{Tr}_B \, \rho_{AB}$). Then:

$$\mathbb{P}[m] = \sum_i \langle i|_A \, P_m^A \rho_A \, |i\rangle = \mathrm{Tr}\left(P_m^A \rho_A\right)$$

This is exactly our old measurement rule!

11.5.1 Relationship between Reduced Density Matrices

Finally, let's think about the relationship between ρ_A and ρ_B.

$$\rho_A = \sum_i \lambda_i \, |\psi_i\rangle\langle\psi_i|$$

However, the ψ_i's are orthogonal and form a basis, so we can combine the terms by $|\psi_i\rangle_A$:

$$|\psi\rangle = \sum_i c_i \, |\psi_i\rangle_A \, |\phi_i\rangle_B$$

where in general, the $|\phi_i\rangle_B$ are not orthogonal. However, we show that in this construction, they actually are (and we can normalize to make them orthonormal)! By definition,

$$\begin{aligned}
\rho_A &= \mathrm{Tr}_B(|\psi\rangle\langle\psi|) \\
&= \sum_k \langle k|_B \left(\sum_{i,j} c_i c_j \, |\psi_i\rangle_A \, |\phi_i\rangle_B \, \langle\psi_j|_A \, \langle\phi_j|_B \right) |k\rangle_B \\
&= \sum_{i,j} c_i c_j \, |\psi_i\rangle_A \, \langle\psi_j|_A \left(\sum_k \langle k|\phi_i\rangle_B \, \langle\phi_j|k\rangle_B \right)
\end{aligned}$$

Note that the k summation is just taking the inner product $\langle\phi_j|\phi_i\rangle_B$.

$$\rho_A = \sum_{i,j} c_i c_j \, |\psi_i\rangle\langle\psi_j|_A \, \langle\phi_j|\phi_i\rangle_B$$

Via the Spectral decomposition, we must have the terms vanish if $i \neq j$, yielding

$$\langle\phi_j|\phi_i\rangle_B = \delta_{ij}$$

Thus the ϕ_i's are orthonormal. In general, the **Schmidt Decomposition** of a state is

$$|\psi\rangle = \sum_i \sqrt{\lambda_i} \, |\psi_i\rangle_A \, |\phi_i\rangle_B$$

where ψ_i and ϕ_i are both simultaneously orthonormal. Thus, we must also have:

$$\rho_B = \sum_i \lambda_i \, |\phi_i\rangle\langle\phi_i|$$

meaning ρ_A and ρ_B have the same eigenvalues!

12 Lecture 12

12.1 Measuring Devices

What is so mysterious about quantum measurements? Well first off, it's weird that the measurement and unitary transformations form two distinct quantum postulates, when they both describe physical processes. Well, we can actually try to describe measurement using unitaries. In fact, any measurement device is itself a quantum system $\mathcal{H}_M$, and a measurement is just an interaction between this device and our measured system $\mathcal{H}_A$. There should be some unitary U_{meas} that describes the measurement. What does this unitary look like? Well, we can see how this acts on basis elements and then on $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, denoting $|??\rangle$ as the state of the device before it's read anything.

$$\begin{aligned}
U_{meas}|0\rangle|??\rangle &= |0\rangle|\text{meas. } 0\rangle \\
U_{meas}|1\rangle|??\rangle &= |1\rangle|\text{meas. } 1\rangle \\
U_{meas}|\psi\rangle|??\rangle &= \alpha|0\rangle|\text{meas. } 0\rangle + \beta|1\rangle|\text{meas. } 1\rangle
\end{aligned}$$

Thus, the measurement device has gotten entangled with the state it's measuring. But, we have a formalism with partial trace for figuring out what state the device is in:

$$\begin{aligned}
\rho_M &= \text{Tr}_A\left[U_{meas}|\psi\rangle|??\rangle\langle??|\langle\psi|U_{meas}^\dagger\right] \\
&= |\alpha|^2|\text{meas. } 0\rangle\langle\text{meas. } 0| + |\beta|^2|\text{meas. } 1\rangle\langle\text{meas. } 1|
\end{aligned}$$

But note, this is the same as an ensemble with $\mathbb{P}[\text{meas. } 0] = |\alpha|^2$ and $\mathbb{P}[\text{meas. } 1] = |\beta|^2$. However, we used the measurement postulate last lecture to get the connection between density matrices and quantum systems. Note that what we've done here is not a derivation of the measurement postulate, but it is a nice way of looking at what quantum measurement is. Similarly, we find for A

$$\rho_A = |\alpha|^2|0\rangle\langle 0| + |\beta|^2|1\rangle\langle 1| \implies \mathbb{P}[|0\rangle] = |\alpha|^2, \mathbb{P}[|1\rangle] = |\beta|^2$$

This means that any future measurements on $\mathcal{H}_A$ that ignore $\mathcal{H}_M$ will agree with this ensemble. Thus, the idea that the wavefunction "collapses" can be explained as unitary evolution with a measurement device.

But what if you don't throw away the measurement device?

$$\begin{aligned}
U_{meas}|+\rangle|??\rangle &= \frac{1}{\sqrt{2}}|0\rangle|\text{meas. } 0\rangle + \frac{1}{\sqrt{2}}|1\rangle|\text{meas. } 1\rangle \\
U_{meas}^\dagger U_{meas}|+\rangle|??\rangle &= |+\rangle|??\rangle
\end{aligned}$$

So, measurement IS reversible! The global state is not an ensemble, but actually a superposition, and thus reversible.

This may seem like it breaks the uncertainty principle, but note that reversing the measurement requires erasing all trace of the outcome of the measurement from the rest of the universe (all entanglements with other systems). So we can't look at it (or if we do, we have to erase our memory).

12.2 Decoherence

In a real setup, very quickly a measurement gets copied (really, entangled) into large quantum systems. For example, if your measurement device has an LED screen, large number of photons have a different state depending on if the measurement device shows a 0 or 1. So, one can think roughly as $|\text{meas. } 0\rangle = |0\rangle|0\rangle\ldots|0\rangle = |0^N\rangle$. In practice $N \in [10^{20}, 10^{30}]$. But, if some "qubit" photon γ leaves our lab, we need to trace over it to figure out our state. This yields:

$$\rho_{S\setminus\{\gamma\}} = |\alpha|^2|0^{N-1}\rangle\langle 0^{N-1}| + |\beta|^2|1^{N-1}\rangle\langle 1^{N-1}|$$

This is a mixed state. We cannot use $U_{meas}^\dagger$ to undo things. This phenomenon is known as **decoherence**.

12.3 Us: Interpretations of Quantum Mechanics

In our class, we distinguish between a *measurement* and an *observation*. A measurement is what we just described: a unitary interaction between a system and a measuring device. An observation is a person actually looking at a measurement outcome. Let's try to discuss this as a unitary transform:

$$U_{obs}(\alpha \left|0\right\rangle \left|\text{meas. } 0\right\rangle + \beta \left|1\right\rangle \left|\text{meas. } 1\right\rangle) \left|hmm\right\rangle = \alpha \left|0\right\rangle \left|\text{meas. } 0\right\rangle \left|\text{sad}\right\rangle + \beta \left|1\right\rangle \left|\text{meas. } 1\right\rangle \left|\text{happy}\right\rangle$$

It's not obvious what it means for my brain to be in a superposition state. Thus, we still need a measurement postulate that superposition is experienced as a probabilistic process weighted by norm squared.

This relates to the interpretations of quantum mechanics. There are two main ones:

1. $|\Psi\rangle$ is *epistemic*. This means it's always fundamentally a quantum version of Bayesian probabilities, meaning that when I see a measurement, I have more information and thus only see one of the terms in my $|\Psi\rangle$. This one is closest to "Copenhagen Interpretation" from 100 years.
2. $|\Psi\rangle$ is *ontic*. $|\Psi\rangle$ is physically real and is everything. This is often considered the many-world interpretation. The probabilities observed in the lab are not accounted for in this model.

12.4 POVM Measurements

If measurements are really interactions with a device, why restrict U_{meas} to only give us $|\text{meas. } 0\rangle$ or $|\text{meas. } 1\rangle$? Let's let it be any unitary. Let's initialize the measurement device in the state $|0\rangle$.

$$U_{meas} \left|\psi\right\rangle \left|0\right\rangle = \sum_m \left|\psi_m\right\rangle \left|m\right\rangle$$

where the $|\psi_m\rangle$ are not normalized. We can rewrite that since this is a linear map, its action on ψ can be seen to be linear:

$$\left|\psi_m\right\rangle = K_m \left|\psi\right\rangle$$

By unitarity,

$$\begin{aligned}
\left\langle 0\right| \left\langle \phi\right| U_{meas}^\dagger U_{meas} \left|\psi\right\rangle \left|0\right\rangle &= \sum_{m_1,m_2} \left\langle m_1\right| \left\langle \phi\right| K_{m_1}^\dagger K_{m_2} \left|\psi\right\rangle \left|m_2\right\rangle \\
&= \sum_{m_1,m_2} \left\langle \phi\right| K_{m_1}^\dagger K_{m_2} \left|\psi\right\rangle \left\langle m_1 | m_2\right\rangle \\
&= \sum_m \left\langle \phi\right| K_m^\dagger K_m \left|\psi\right\rangle \\
\left\langle \phi | \psi\right\rangle &= \left\langle \phi\right| \left(\sum_m K_m^\dagger K_m\right) \left|\psi\right\rangle
\end{aligned}$$

For this to be true for any $|\phi\rangle$ and $|\psi\rangle$, we must have

$$\sum_m K_m^\dagger K_m = I$$

The K_m's are usually known as **Kraus operators**. Now to get a measurement, we just find the probability according to the measurement postulate

$$\begin{aligned}
\mathbb{P}\left[m\right] &= \left|\left\langle m\right| U_{meas} \left|\psi\right\rangle \left|0\right\rangle\right|^2 \\
&= \sum_{m_1,m_2} \left\langle m_1\right| \left\langle \psi\right| K_{m_1}^\dagger \left|m\right\rangle \left\langle m\right| K_{m_2} \left|\psi\right\rangle \left|m_2\right\rangle \\
&= \sum_{m_1,m_2} \left\langle \psi\right| K_{m_1}^\dagger \left\langle m_1 | m\right\rangle \left\langle m | m_2\right\rangle K_{m_2} \left|\psi\right\rangle \\
&= \left\langle \psi \middle| K_m^\dagger K_m \middle| \psi\right\rangle = \left\langle K_m^\dagger K_m\right\rangle
\end{aligned}$$

So note, that $\Pi_m := K_m^\dagger K_m$ is all the matters. Note that an operator is in the form $A^\dagger A$ if and only if it is positive semi-definite. Thus we have:

$$\sum_m \Pi_m = I, \Pi_m \geq 0$$

Thus, there is no reason that these Π_m's need to be projection operators at all. They just need to be PSD. If they were projection operators, we'd get the **projection-valued measure (PVM)** from last time. Instead these are the **positive-operator-valued measure (POVM)**. It turns out, this lets you learn about weak or noisy measurements. Consider the two operators

$$\Pi_0 = \frac{3}{4} |0\rangle\langle 0| + \frac{1}{4} |1\rangle\langle 1|, \Pi_1 = \frac{1}{4} |0\rangle\langle 0| + \frac{3}{4} |1\rangle\langle 1|$$

These clearly add up to the identity. Then for the original state $|0\rangle$, the device enters a state where:

$$\mathbb{P}[0] = \langle 0|\Pi_0|0\rangle = 3/4, \mathbb{P}[1] = 1/4$$

and if we instead started with a $|1\rangle$,

$$\mathbb{P}[0] = \langle 0|\Pi_1|0\rangle = 1/4, \mathbb{P}[1] = 3/4$$

This is a noisy measurement because $3/4$ of the time, we get the correct answer and $1/4$ of the time, we get something completely wrong.

Remember, $|\psi\rangle \to K_m |\psi\rangle$ when $K_m^\dagger K_m = \Pi_m$; so a particularly nice Kraus operator is $\Pi_m^{1/2}$. Taking our previous example, we have:

$$K_0 = \frac{\sqrt{3}}{2} |0\rangle\langle 0| + \frac{1}{2} |1\rangle\langle 1|, \qquad K_1 = \frac{1}{2} |0\rangle\langle 0| + \frac{\sqrt{3}}{2} |1\rangle\langle 1|$$

$$K_0 |+\rangle = \frac{\sqrt{3}}{2} |0\rangle + \frac{1}{2} |1\rangle, \qquad K_1 |+\rangle = \frac{1}{\sqrt{2}} |0\rangle + \frac{\sqrt{3}}{2} |1\rangle$$

13 Lecture 13

13.1 Shannon Entropy

Suppose we have some (discrete) classical probability distribution $\{p_i\}_{i=1}^n$. We'd like to somehow characterize the amount of uncertainty in this distribution. One natural notion would be to count the number of outcomes which have probability that is nonzero. However, this notion does not acknowledge a difference between all the p_i's being equal and the case where $p_1 > 0.99$ and the rest are tiny. The former definitely has more uncertainty than the latter. The better notion is **Shannon Entropy**, which we will motivate before introducing.

One should take N independent samples and take the limit as $N \to \infty$. By the (weak) law of large numbers with probability tends to 1, the fraction of outcome i, $N_i/N \to p_i$. We define a "typical outcome" to be a sample where this fraction is true (up to some ϵ wiggle room). The number of typical outcomes, by combinatorics:

$$N_{typ} = \sum_{\text{typical } \{N_i\}} \frac{N!}{\prod_i N_i!}$$

The number of typical $\{N_i\}$ grows at most polynomially with N, because for a typical N_i, $N_i/N \in p_i \pm \epsilon$, so $N_i \in Np_i \pm N\epsilon$, so there are roughly $(2N\epsilon)^n$ choices, which is still polynomial (n is the number of possible outcomes, remember). However, the probability of the factorials is going to turn out to be exponential, so we will ignore the summation.

$$\begin{aligned}
\log N_{typ} &\approx \log(N!) - \sum_i \log(N_i!) \\
&\approx N \log N - \sum_i N_i \log N_i \\
&= \sum_i N_i \log N - \sum_i N_i \log N_i \\
&= -\sum_i N_i \log\left(\frac{N_i}{N}\right) \\
&= -N \sum_i p_i \log p_i
\end{aligned}$$

This motivates the following definition.

Definition 13.1
The Shannon Entropy of a probability distribution $\mathbf{p} = \{p_i\}_{i=1}^N$ is

$$H(\mathbf{p}) = -\sum_i p_i \log p_i$$

Intuitively, the Shannon entropy denotes how many bits are required to communicate a typical sample from the distribution. From another direction, we can think of the probability of getting a typical outcome.

$$\begin{aligned}
p_{typ} &= \prod_i p_i^{N_i} \\
\log p_{typ} &= \sum_i N_i \log p_i \\
&\approx -NH(\mathbf{p})
\end{aligned}$$

But then, we take

$$\begin{aligned}\log p_{typ}N_{typ} &= \log p_{typ} + \log N_{typ} \\ &\approx -NH(\mathbf{p}) + NH(\mathbf{p}) \\ &= 0\end{aligned}$$

And thus we have

$$p_{typ}N_{typ} \approx 1$$

As a sanity check, this makes sense because the sum of the probabilities of all of the typical outcomes should be 1. This gives us the sketch of the Shannon source-coding theorem.

Theorem 13.1 (Source-coding)
It requires $N \cdot H(\mathbf{p}) + o(N)$ bits to communicate a typical outcome of N samples from $\{p_i\}$.

Thus, one can say that it takes amortized $H(\mathbf{p})$ bits of information per sample to communicate a typical string.

13.2 Entanglement Entropy

Let's extend our analysis to quantum. Take $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and suppose Alice and Bob share a state $|\Psi\rangle^{\otimes N}$ and Alice wants to send her part of $|\Psi\rangle^{\otimes N}$ to Charlie so that with high probability a state shared by Charlie and Bob becomes $|\tilde{\Psi}\rangle_{(N)} \approx |\Psi\rangle^{\otimes N}$.

If the initial state is a product state $|\Psi\rangle = |\psi\rangle_A |\phi\rangle_B$, then 0 qubits are needed. Charlie can prepare $|\psi\rangle^{\otimes N}$ in his lab and Bob can contribute his state.

Recall that we previously derived

$$\begin{aligned}|\Psi\rangle &= \sum_i \sqrt{\lambda_i} |\psi_i\rangle_A |\phi_i\rangle_B \\ \rho_A &= \sum_i \lambda_i |\psi_i\rangle\langle\psi_i| \\ \rho_B &= \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|\end{aligned}$$

We define the typical subspace $\mathcal{H}_{typ} \subseteq \mathcal{H}_A$ as

$$\mathcal{H}_{typ} = \text{span}\left\{ |\psi_{i_1}\rangle |\psi_{i_2}\rangle \dots |\psi_{i_N}\rangle : \frac{N_i}{N} \approx \lambda_i \right\}$$

We define P_{typ} as the projector onto $\mathcal{H}_{typ}$. Applying the projector,

$$\langle\Psi|^{\otimes N} P_{typ} |\Psi\rangle^{\otimes N} = \sum_{\text{typical } \{i_j\}} \prod_{j=1}^{N} \lambda_{i_j} = \sum_i p_i \approx 1$$

The dimension of the typical subspace can be recovered by noting that it's just the number of typical strings, which we already found was:

$$\log \dim(\mathcal{H}_{typ}) \approx NH(\lambda)$$

Now we can define the protocol. Alice measures whether $|\Psi\rangle^{\otimes N}$ are in $\mathcal{H}_{typ}$ The new state becomes:

$$|\tilde{\Psi}_{(N)}\rangle = \frac{P_{typ} |\Psi\rangle^{\otimes N}}{\sqrt{\langle\Psi|^{\otimes N} P_{typ} |\Psi\rangle^{\otimes N}}} \approx |\Psi\rangle^{\otimes N}$$

We can now send this to Charlie using $NH(\lambda)$ qubits and we're done describing the state. This motivates the following quantum entropy definition (taking operator log by its infinite series):

Definition 13.2
The Von Neumann Entropy of a density matrix ρ is:

$$S(\rho) = -\operatorname{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log \lambda_i$$

If we have a pure state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, then $S(\rho_A) = S(\rho_B) = H\left(\vec{\lambda}\right)$, and we term this the **entanglement entropy**.

13.3 Quantum Channels

We've already talked about noisy states and noisy measurement. How can we define a noisy quantum evolution? It turns out it can be described by a few things. Start with some state $\rho_A \in \mathcal{H}_A$.

1. Take a fixed state $|0\rangle \in \mathcal{H}_X$ and add it to ρ_A to obtain $\rho_A \otimes |0\rangle\langle 0|_X$
2. Take an arbitrary unitary evolution U_{AX} and use it to give new state $U_{AX}(\rho_A \otimes |0\rangle\langle 0|_X)U_{AX}^\dagger$.
3. Write the entire Hilbert space as $\mathcal{H} \simeq \mathcal{H}_A \otimes \mathcal{H}_X \simeq \mathcal{H}_B \otimes \mathcal{H}_E$. Then throw away $\mathcal{H}_E$, yielding

$$\rho_B = \operatorname{tr}_E\left[U_{AX}(\rho_A \otimes |0\rangle\langle 0|_X)U_{AX}^\dagger\right] = \mathcal{N}(\rho_A)$$

We call $\mathcal{N}$, which takes an operator and spits out an operator, a **superoperator**.

We can also describe the first two steps as applying $V = U_{AX}|0\rangle_X$, where $V : \mathcal{H}_A \to \mathcal{H}$ and is almost unitary, namely:

$$V^\dagger V = I_A, VV^\dagger \neq I_B \otimes I_E$$

So, the map V preserves inner product and is termed an **isometry**. Our channel thus becomes:

$$\mathcal{N}(\rho) = \operatorname{Tr}_E(V\rho V^\dagger)$$

In particular, this is called the **Stinespring representation** of the channel. Furthermore, we can also write this with Kraus operators:

$$V|\Psi\rangle = \sum_i K_i |\Psi\rangle |i\rangle_E$$

where $V^\dagger V = I$ implies $\sum_i K_i^\dagger K_i = I$. The channel then becomes

$$\mathcal{N}(\rho) = \operatorname{Tr}_E\left(\sum_{i,j} K_i \rho K_j^\dagger \otimes |i\rangle\langle j|\right) = \sum_i K_i \rho K_i^\dagger$$

This is called the **Kraus representation** of the channel.

Let's consider a few examples of channels.

Example 13.1 (Basic Examples)

1. The Dephasing Channel.

$$K_0 = \sqrt{1 - p/2}I, \quad K_1 = \sqrt{p/2}Z$$

Then the channel is

$$\mathcal{N}(\rho) = \left(1 - \frac{p}{2}\right)\rho + \frac{p}{2}Z\rho Z^\dagger$$

This means that with probability $p/2$, the phase gets flipped, else we do not.

$$\rho = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies \mathcal{N}(\rho) = \begin{pmatrix} a & (1-p)b \\ (1-p)c & d \end{pmatrix}$$

If $p = 1$, then

$$\mathcal{N}(\rho) = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$$

Then, we get a mixture of $|0\rangle$ and $|1\rangle$; all phase information is lost.

2. The Depolarizing Channel.

$$K_0 = \sqrt{1 - \frac{3p}{4}} I, \quad K_1 = \sqrt{\frac{p}{4}} X, \quad K_2 = \sqrt{\frac{p}{4}} Y, \quad K_3 = \sqrt{\frac{p}{4}} Z$$

Then the channel is

$$\mathcal{N}(\rho) = \left(1 - \frac{3p}{4}\right)\rho + \frac{p}{4} X\rho X + \frac{p}{4} Y\rho Y + \frac{p}{4} Z\rho Z$$

which can be rewritten as

$$\mathcal{N}(\rho) = (1-p)\rho + \frac{p}{2} I$$

which means as $p \to 1$, the state is less like ρ and becomes closer to maximally mixed.

13.4 Properties of Quantum Channels

It turns out quantum channels are really pretty general objects. Here are some properties they satisfy:

1. A quantum channel is linear.
$$\mathcal{N}(\alpha\rho + \alpha\sigma) = \alpha\mathcal{N}(\rho) + \beta\mathcal{N}(\sigma)$$
2. A quantum channel is trace-preserving.
$$\operatorname{Tr}\mathcal{N}(\rho) = \operatorname{Tr}\rho$$
3. A quantum channel is positive. If $\rho \geq 0$, then
$$\mathcal{N}(\rho) = \sum_i K_i \rho K_i^\dagger \geq 0$$

Note that 2 and 3 combined say that $\mathcal{N}$ maps density matrices to density matrices!

Is any superoperator satifying these properties a quantum channel? The answer is no. A simple counterexample is the map that takes $\rho \mapsto \rho^T$, i.e. the transpose map. Consider a state $|\Phi^+\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ of two qubits. The density matrix of this state is:

$$\rho_{AB} = \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| = \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|)$$

We will apply the transpose to the first qubit.

$$\rho_{AB}^{T_A} = \frac{1}{2}(|00\rangle\langle 00| + |10\rangle\langle 01| + |01\rangle\langle 10| + |11\rangle\langle 11|)$$

But this is not a density matrix at all. The state $|01\rangle - |10\rangle$ has eigenvalue $-1/2$. So we need a fourth condition:

4. A quantum channel is completely positive.
$$\rho_{AR} \geq 0 \implies (\mathcal{N} \otimes \mathrm{Id}_R)(\rho_{AR}) \geq 0$$

Note that 4 implies 3. In fact, we will see that next lecture, any completely positive, trace-preserving, linear superoperator is a quantum channel.

14 Lecture 14

14.1 Quantum Channel Generality

We begin by recalling the quantum channels, and state that these are actually very general objects.

Theorem 14.1
A superoperator $\mathcal{N}$ is completely positive, trace-preserving, and linear *if and only if* it is a quantum channel.
Proof
We have already shown the if step. To show the only if, consider $\mathcal{H}_R \cong \mathcal{H}_A$, where the $\mathcal{H}_R$ space is the one which will be acted on by the identity and $\mathcal{H}_A$ is the actual input space. Consider $\{|j\rangle\}$ to be the orthonormal basis for this space.
Consider $|\Phi\rangle = \sum_j |j\rangle_A |j\rangle_R$. Now, applying the superoperator tensored with the identity superoperator:

$$(\mathcal{N} \otimes \mathrm{Id}_R)(|\Phi\rangle\langle\Phi|) = \sum_i p_i |\Phi_i\rangle\langle\Phi_i|$$

for $|\Phi_i\rangle \in \mathcal{H}_B \otimes \mathcal{H}_R$. Because of positive semidefiniteness, we can take the square root of p_i and write $|\Phi_i\rangle$ as a linear combination of its basis elements:

$$\sqrt{p_i} |\Phi_i\rangle = \sum_{\alpha,j} K_i^{\alpha j} |\alpha\rangle_B |j\rangle_R$$

Now, we define the map

$$K_i = \sum_{\alpha,j} K_i^{\alpha j} |\alpha\rangle\langle j|$$

Then, this means $\sqrt{p_i} |\Phi_i\rangle = K_i |\Phi\rangle$ (we contract over exactly the coefficients of $|\Phi_i\rangle$), so the output of the channel tensored with the identity yields

$$\sum_i p_i |\Phi_i\rangle\langle\Phi_i| = \sum_i K_i |\Phi\rangle\langle\Phi| K_i^\dagger$$

Now consider another state $|\psi\rangle = \sum_j c_j |j\rangle \in \mathcal{H}_A$ and define $|\overline{\psi}\rangle = \sum_j c_j^* |j\rangle \in \mathcal{H}_R$. Then:

$$\langle\overline{\psi}|\Phi\rangle = \sum_{j',j} c'_j \langle j'|j\rangle_R |j\rangle_A = \sum_j c_j |j\rangle = |\psi\rangle$$

and thus we have

$$\langle\overline{\psi}| (\mathcal{N} \otimes \mathrm{Id}_R)(|\Phi\rangle\langle\Phi|) |\overline{\psi}\rangle = \sum_i K_i \langle\overline{\psi}| |\Phi\rangle\langle\Phi| |\overline{\psi}\rangle K_i^\dagger$$

$$\mathcal{N}(\langle\overline{\psi}| |\Phi\rangle\langle\Phi| |\overline{\psi}\rangle) = \sum_i K_i |\psi\rangle\langle\psi| K_i^\dagger$$

$$\mathcal{N}(|\psi\rangle\langle\psi|) = \sum_i K_i |\psi\rangle\langle\psi| K_i^\dagger$$

Now, by a simple superposition, this works for any density operator $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$

$$\begin{aligned}
\mathcal{N}(\rho) &= \sum_j \lambda_j \mathcal{N}(|\psi_j\rangle\langle\psi_j|) \\
&= \sum_{i,j} \lambda_j K_i |\psi_j\rangle\langle\psi_j| K_i^\dagger \\
&= \sum_i K_i \left(\sum_j \lambda_j |\psi_j\rangle\langle\psi_j| \right) K_i^\dagger \\
&= \sum_i \lambda_i K_i \rho K_i^\dagger
\end{aligned}$$

Finally, we will show that the Kraus operators are properly normalized using the trace-preserving property of the superoperator:

$$\operatorname{Tr}\mathcal{N}(\rho) = \sum_i \operatorname{Tr}\left(K_i \rho K_i^\dagger\right)$$

$$\operatorname{Tr}\rho = \operatorname{Tr}\left(\rho\left(\sum_i K_i^\dagger K_i\right)\right)$$

Take $\rho = |\psi\rangle\langle\psi|$. Then the above property implies

$$\left\langle\psi\middle|\sum_i K_i^\dagger K_i\middle|\psi\right\rangle = \langle\psi|\psi\rangle \implies \sum_i K_i^\dagger K_i = I$$

Thus, $\mathcal{N}$ is a valid quantum channel (in Kraus representation).

14.2 Master Equations

Recall that we could define discrete noiseless evolution by a unitary operator

$$\rho \mapsto U\rho U^\dagger$$

Furthermore, we could define continuous noiseless evolution by the Von Neumann equation (the density operator version of the Schrodinger equation)

$$i\frac{\mathrm{d}\rho}{\mathrm{d}t} = [H, \rho]$$

Now, for a noisy (or open-system) discrete evolution, we now have quantum channels

$$\rho \mapsto \mathcal{N}(\rho)$$

Now, we expect some kind of open-system continuous evolution of the form $\frac{\mathrm{d}\rho_A}{\mathrm{d}t} = f(\rho_A)$. Unfortunately, this is not usually possible to do, because the evolution is not dependent on purely on initial condition; information can leave the A system and then come back later. Call $\mathcal{H}_A \cong \mathcal{H}_E$, where the E system is the environment. At $t = 0$, suppose the state is $|\psi\rangle_A |0\rangle_E$. We will write

$$U(t_0) = e^{-iHt_0} = \text{SWAP}, \quad U(t_0)|i\rangle|j\rangle = |j\rangle|i\rangle$$

So at $t = t_0$, $\rho_A(t_0) = |0\rangle\langle 0|$. All the information has escaped out of A. But afterwards, $\rho_A(2t_0) = |\psi\rangle\langle\psi|$. The information flowed back into A. At time $2t_0$ we needed the state at 0, so the system is not memoryless and we cannot create an evolution law in the manner we described.

However, if $|\mathcal{H}_E| \gg |\mathcal{H}_A|$ and the environment is evolving sufficiently "fast," then information that escapes the system and doesn't immediately return is gone forever to the environment. To model this, we replace $\mathcal{H}_E$ by some "clean copy" in a fixed state after every $\Delta t_{forget} \ll 1$ timesteps. If we are fine with coarse-grained evolution, then we will look at updating the state in increments δt where $\Delta t_{forget} \ll \delta t \ll \Delta t_{system}$, where the last term is the time for the system to evolve significantly. We will figure this as

$$\mathcal{N}_{\delta t}(\rho(t)) = \rho(t + \delta t) = \rho(t) + O(\delta t)$$

By Taylor expansion:

$$\mathcal{N}_{\delta t} = \mathrm{Id} + \delta t\mathcal{L}$$

where the superoperator $\mathcal{L}$ is called the **Lindbladian** of the system. By rearranging $\rho(t + \delta t) = \rho(t) + \delta t\mathcal{L}(\rho)$ and taking limits, we have

$$\implies \frac{\mathrm{d}\rho}{\mathrm{d}t} = \mathcal{L}(\rho)$$

Assuming that $\mathcal{L}$ is independent of time, then the solution to the equation is

$$\rho(t) = \lim_{n\to\infty} \left(\mathrm{Id} + \frac{t}{n}\mathcal{L}\right)^n (\rho(0))$$
$$\rho(t) = e^{t\mathcal{L}}(\rho(0))$$

But what does the Lindbladian really look like? Note that $\mathcal{N}_{\delta t}$ is a quantum channel, so it has a Kraus representation:

$$\mathcal{N}(\rho(t)) = \sum_i K_i \rho K_i^\dagger = \rho + O(\delta t)$$

The easiest way to write this is $K_0 = I + \delta t(-iH + K)$ where H and K are Hermitian, resulting in a split into anti-Hermitian ($-iH$) and Hermitian (K) parts. Furthermore, K_i cannot have any constant terms: they're all accounted for. To get a linear term in $K_i \rho K_i^\dagger$, we need order 1/2 terms for $i \neq 0$: $K_i = \sqrt{\delta t} L_i$. But by normalization:

$$\sum_i K_i^\dagger K_i = I = I + \delta t(iH + K) + \delta t(-iH + K) + \sum_{i>0} L_i^\dagger L_i \delta t$$
$$0 = 2\delta t K + \delta t \sum_{i>0} L_i^\dagger L_i$$
$$K = -\frac{1}{2}\sum_{i>0} L_i^\dagger L_i$$

and thus, we have

$$\begin{aligned}\frac{\mathrm{d}\rho}{\mathrm{d}t} &= \frac{\mathcal{N}_{\delta t}(\rho) - \rho}{\delta t} \\ &= \frac{1}{\delta t}\left[\sum_i K_i \rho K_i^\dagger - \rho\right] \\ &= \left(-iH - \frac{1}{2}\sum_{i>0} L_i^\dagger L_i\right)\rho + \rho\left(iH - \frac{1}{2}\sum_{i>0} L_i^\dagger L_i\right) + \sum_i L_i \rho L_i^\dagger\end{aligned}$$

Let's clean this up.

Theorem 14.2
Continuous-time noisy evolution of a state, where each step is quantum channel, is described by the **Lindblad master equation**:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -i[H, \rho] + \sum_{i>0}\left(L_i \rho L_i^\dagger - \frac{1}{2}L_i^\dagger L_i \rho - \frac{1}{2}\rho L_i^\dagger L_i\right)$$

The first term is what one gets from normal Unitary evolution from a Hamiltonian H. The second term $\sum_{i>0}\left(L_i \rho L_i^\dagger\right)$ is the effects of noise, where L_i are considered "quantum jump" operators where the jump is described by $|\psi\rangle \mapsto L_i |\psi\rangle$. Finally, the last two terms $\sum_{i>0} \frac{1}{2} L_i^\dagger L_i \rho$ and $\sum_{i>0} \rho L_i^\dagger L_i$ preserve the normalization of ρ overall.

This was a lot of formalism, let's explore an example.

Example 14.1 (Photon Emission by Excited Electron State)
Consider a photon emission by an excited electron state. We will model our A system as the electron and the photon as noise. We model $|0\rangle$ as the ground state of the electron and $|1\rangle$ as the excited state of the electron. We say the eigenvalue of the energy level $|0\rangle$ is 0 and $|1\rangle$ is E, giving the Hamiltonian as $H = E\,|1\rangle\langle 1|$. We will model $L_1 = \sqrt{\Gamma}\,|0\rangle\langle 1|$, where Γ is the photon emission rate. Now, let's try to solve the master equation.

First, we express the state ρ explicitly as follows:

$$\rho = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} = \rho_{00} |0\rangle\langle 0| + \rho_{01} |0\rangle\langle 1| + \rho_{10} |1\rangle\langle 0| + \rho_{11} |1\rangle\langle 1|$$

Then note that the first and last terms commute with the Hamiltonian and one of the cross terms is:

$$\begin{aligned} H\rho &= E |1\rangle\langle 1| \rho_{01} |0\rangle\langle 1| = 0 \\ \rho H &= \rho_{01} |0\rangle\langle 1| E |1\rangle\langle 1| = E\rho_{01} |0\rangle\langle 1| \end{aligned}$$

the other cross term will be negative of this, so we have

$$\begin{aligned} -i[H, \rho] &= -i(H\rho - \rho H) \\ &= \begin{pmatrix} 0 & iE\rho_{01} \\ -iE\rho_{10} & 0 \end{pmatrix} \end{aligned}$$

Now, the other terms:

$$\begin{aligned} L_1 \rho L_1^\dagger &= \Gamma |0\rangle\langle 1| \rho |1\rangle\langle 0| = \Gamma\rho_{11} |0\rangle\langle 0| \\ -\frac{1}{2} L_1^\dagger L_1 \rho &= -\frac{\Gamma}{2} |1\rangle\langle 1| \rho = -\frac{\Gamma}{2}(\rho_{10} |1\rangle\langle 0| + \rho_{11} |1\rangle\langle 1|) \\ -\frac{1}{2} \rho L_1^\dagger L_1 &= -\frac{\Gamma}{2} \rho |1\rangle\langle 1| = -\frac{\Gamma}{2}(\rho_{11} |1\rangle\langle 1| + \rho_{01} |0\rangle\langle 1|) \end{aligned}$$

Putting it together,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \begin{pmatrix} \Gamma\rho_{11} & iE\rho_{01} - \frac{1}{2}\Gamma\rho_{01} \\ -iE\rho_{10} - \frac{1}{2}\Gamma\rho_{10} & -\Gamma\rho_{11} \end{pmatrix} \implies \rho(t) = \begin{pmatrix} \rho_{00}(0) + \left(1 - e^{-\Gamma t}\right)\rho_{11}(0) & e^{\left(iE - \frac{\Gamma}{2}\right)t}\rho_{01}(0) \\ e^{\left(-iE - \frac{\Gamma}{2}\right)t}\rho_{10}(0) & e^{-\Gamma t}\rho_{11}(0) \end{pmatrix}$$

At any time, it can drop down, so as $t \to \infty$, the density approaches $|0\rangle\langle 0|$. The cross-terms gain a phase based on the eigenenergy, which is what that Hamiltonian would do on its own.

15 Lecture 15

15.1 Quantum Error-Correction: Intro

A long time ago, in the 1950s, there were analog computers. Rather than storing data as 0s and 1s, you store it as real numbers (these can be mechanical systems or electrical systems). A noiseless analog computer is very powerful; adding up arbitrary precision real numbers is very powerful. However, in the real world, these are noisy, which actually makes them computationally equivalent to digital computers. And later we realized that digital computers are far more scalable than analog ones, so that's why our computer systems are digital today!

Back when quantum computing was just starting up, people would scared it would be analog computing all over again. Peter Shor invented the Shor code to fix this problem.

15.2 Classical Error-Correction

We have a bitstring with characters in $\{0, 1\}$ sitting in our hard drive. Suppose that cosmic ray comes in and spontaneously flips some bits. To protect against this, the simplest thing to do is the **repetition code**, e.g. storing the logical bit 0 as 000 and the logical bit 1 as 111. This redundancy allows us to protect against some errors. Suppose you see 101; we would use *majority rule* and declare the bit was 1, encoding as 111. The probability that it was initially 000 is small (specifically $\frac{1}{8}$).

What's the probability that we get this wrong? If the probability of any individual bit getting flipped is $p \ll 1$ (independently), the probability of any bit wrong is $O(p^2)$, which is very small.

15.3 Shor Code

The quantum error-correction problem seems impossible because there are infinitely many errors; we will later show that it actually is possible using the **Shor Code**.

For now, let us assume that the only possible error that could occur on a qubit is an X gate (the quantum bit-flip). For a single bit state we'd like to encode a logical qubit as $\left|\tilde{\psi}\right\rangle \mapsto |\psi\rangle |\psi\rangle |\psi\rangle$, but this is impossible due to no-cloning. We have to settle for the CNOT copying we saw prior:

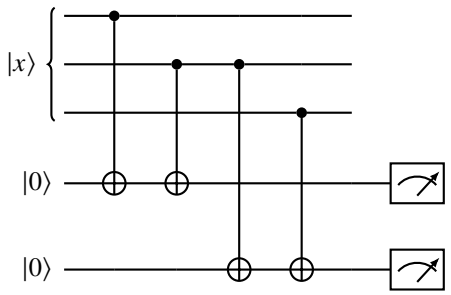
$$\left|\tilde{0}\right\rangle \mapsto |000\rangle, \left|\tilde{1}\right\rangle \mapsto |111\rangle$$

And we can encode a general state with linearity

$$\left|\tilde{\psi}\right\rangle = \alpha \left|\tilde{0}\right\rangle + \beta \left|\tilde{1}\right\rangle \mapsto \alpha |000\rangle + \beta |111\rangle$$

We say that we want to correct at most one bitflip X (the other qubits have the identity applied to them).

After noise, we want to correct the error using the same "majority rule algorithm." If we had a nonsuperpositional computational state, then we could measure in computational basis and be done. The problem is we cannot measure without destroying information. So instead, we build the following circuit, calling our noisy codeword $|x\rangle = |x_1, x_2, x_3\rangle$.



This maps a computational basis state

$$|x_1, x_2, x_3\rangle |00\rangle \mapsto |x_1, x_2, x_3\rangle |x_1 \oplus x_2, x_2 \oplus x_3\rangle$$

and then measures the ancilla bits. Let us analyze the cases for the resulting state:

Case 1. If there was no error, we have $|00\rangle$ on the ancilla (extra) bits and measure 00. The state collapses back to $\alpha |000\rangle + \beta |111\rangle$.

Case 2. If there was a X_1 error, we have $|10\rangle$ on the ancilla bits and measure 10. The state collapses to $\alpha |100\rangle + \beta |011\rangle$.

Case 3. If there was a X_2 error, we have $|11\rangle$ on the ancilla bits and measure 11. The state collapses to $\alpha |010\rangle + \beta |101\rangle$.

Case 4. If there was an X_3 error, we have $|01\rangle$ on the ancilla bits and measure 01. The state collapses to $\alpha |001\rangle + \beta |110\rangle$.

In all cases, it's clear what's left to do, just apply X_i to the bit i if it was flipped, otherwise don't do anything. The measurement on the ancilla bits is called a **Syndrome Measurement**. Recall the phase flip operator

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

Applying the syndrome measurement e.g. measuring s on the first ancilla qubit is the same as measuring $(-1)^s$ in the basis $Z_1 Z_2$. To see this, note that measuring with Z_1 on state $|x_1, x_2, x_3\rangle$ will yield $(-1)^{x_1}$. Thus, multiplying the two will XOR the bits, yielding $Z_1 Z_2 = (-1)^{x_1 \oplus x_2}$. We can then summarize:

Error	Ancilla measurement	Syndrome measurement	Error correction
I	00	$Z_1 Z_2 = 1, Z_2 Z_3 = 1$	I
X_1	10	$Z_1 Z_2 = -1, Z_2 Z_3 = 1$	X_1
X_2	11	$Z_1 Z_2 = -1, Z_2 Z_3 = -1$	X_2
X_3	01	$Z_1 Z_2 = 1, Z_2 Z_3 = -1$	X_3

But X gates are not the only things that could happen to qubits. What if a Z gate was sneakily applied instead?

$$Z_1(\alpha |000\rangle + \beta |111\rangle) = \alpha |000\rangle - \beta |111\rangle = \tilde{Z}\left(\alpha \left|\tilde{0}\right\rangle + \beta \left|\tilde{1}\right\rangle\right)$$

There's no way to distinguish an error from a different logical state. We need to find a different code. But there is symmetry between X and Z: they are the same under a change of basis. So, we can then define the repetition code as:

$$|\tilde{+}\rangle \mapsto |+++\rangle, \quad |\tilde{-}\rangle \mapsto |---\rangle$$

Because we have this duality between them, we get the same table

Error	Ancilla measurement	Syndrome measurement	Error correction
I	++	$X_1 X_2 = 1, X_2 X_3 = 1$	I
X_1	−+	$X_1 X_2 = -1, X_2 X_3 = 1$	Z_1
X_2	−−	$X_1 X_2 = -1, X_2 X_3 = -1$	Z_2
X_3	+−	$X_1 X_2 = 1, X_2 X_3 = -1$	Z_3

But now by the same logic, we cannot correct bit flip errors. To fix this, we can now perform code concatenation, i.e. encode our logical bits with the phase flip code and then the bit flip code. This yields a 9-qubit code:

$$\left|\tilde{\tilde{+}}\right\rangle \mapsto |\tilde{+}\rangle |\tilde{+}\rangle |\tilde{+}\rangle \mapsto \frac{1}{2\sqrt{2}}(|000\rangle + |111\rangle)^{\otimes 3}$$

$$\left|\tilde{\tilde{-}}\right\rangle \mapsto |\tilde{-}\rangle |\tilde{-}\rangle |\tilde{-}\rangle \mapsto \frac{1}{2\sqrt{2}}(|000\rangle - |111\rangle)^{\otimes 3}$$

On decoding, we first error-correct within each bit-flip code (i.e. $|x_1,x_2,x_3\rangle$, $|x_4,x_5,x_6\rangle$, $|x_7,x_8,x_9\rangle$):

$$Z_1Z_2, Z_2Z_3, Z_4Z_5, Z_5Z_6, Z_7Z_8, Z_8Z_9$$

nd apply X gates as necessary. What about phase-flip errors? Notice that

$$Z_1\left|\tilde{\tilde{\psi}}\right\rangle = Z_2\left|\tilde{\tilde{\psi}}\right\rangle = Z_3\left|\tilde{\tilde{\psi}}\right\rangle = \tilde{Z}_{\tilde{1}}\left|\tilde{\tilde{\psi}}\right\rangle$$

This is similar for $Z_{4,5,6}$ and $Z_{7,8,9}$ with $\tilde{Z}_{\tilde{2}}$ and $\tilde{Z}_{\tilde{3}}$. This means that to correct for phase flip errors in the first bit-flip code, we just have to measure $\tilde{X}_{\tilde{1}}\tilde{X}_{\tilde{2}}$. What do these operators actually mean as physical operators? Well, they flip the entire logical bit:

$$\tilde{X}_{\tilde{1}}\left|\tilde{0}\tilde{0}\tilde{0}\right\rangle = \left|\tilde{1}\tilde{0}\tilde{0}\right\rangle$$
$$\tilde{X}_{\tilde{1}}\left|000000000\right\rangle = \left|111000000\right\rangle$$

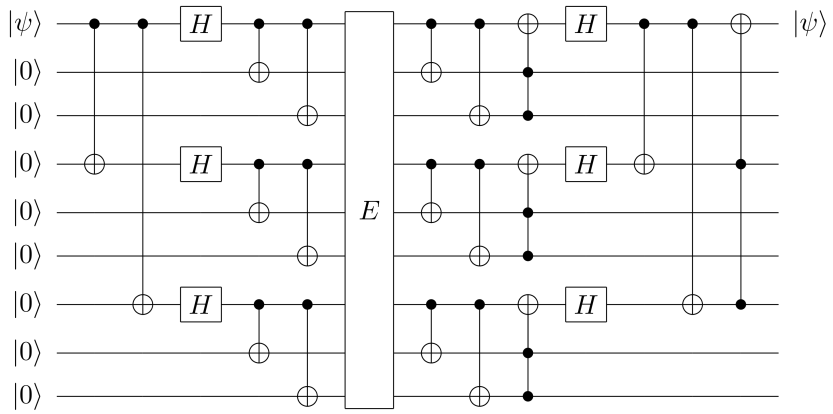
Thus, we have $\tilde{X}_{\tilde{1}} = X_1X_2X_3$, $\tilde{X}_{\tilde{2}} = X_4X_5X_6$, and $\tilde{X}_{\tilde{3}} = X_7X_8X_9$. So on the outer code, we are measuring

$$\tilde{X}_{\tilde{1}}\tilde{X}_{\tilde{2}} = X_1X_2X_3X_4X_5X_6$$
$$\tilde{X}_{\tilde{2}}\tilde{X}_{\tilde{3}} = X_4X_5X_6X_7X_8X_9$$

Hence, the summary table for phase-flip error correction is

Error	Syndrome measurement		Error correction
I	$X_1X_2X_3X_4X_5X_6 = 1$,	$X_4X_5X_6X_7X_8X_9 = 1$	I
$Z_1Z_2Z_3$	$X_1X_2X_3X_4X_5X_6 = -1$,	$X_4X_5X_6X_7X_8X_9 = 1$	Z_1
$Z_4Z_5Z_6$	$X_1X_2X_3X_4X_5X_6 = -1$,	$X_4X_5X_6X_7X_8X_9 = -1$	Z_4
$Z_7Z_8Z_9$	$X_1X_2X_3X_4X_5X_6 = 1$,	$X_4X_5X_6X_7X_8X_9 = -1$	Z_7

Thus, we have an quantum error correcting code that can correct any 1-qubit bit or phase flip. This is known as the 9-qubit **Shor code**, depicted by the circuit below:



where E is where any bit or phase flip error may occur. Using the Shor code, we can in fact correct all the pauli errors, e.g. consider $Y = iXZ$, and the i is irrelevant (we will drop the global phase for simplicity) so

$$Y\left|\tilde{\tilde{\psi}}\right\rangle = X_k\tilde{Z}_k\left|\tilde{\tilde{\psi}}\right\rangle \mapsto \tilde{Z}_k\left|\tilde{\tilde{\psi}}\right\rangle \mapsto \left|\tilde{\tilde{\psi}}\right\rangle$$

This is due to how we can correct errors one at a time.

15.4 General Quantum Errors

It turns out the Shor code can error-correct any general error introduced by a quantum channel. Recall the Kraus operator representation

$$\mathcal{N}(\rho) = \sum_i K_i \rho K_i^\dagger$$

We will consider only a single state, because if we can correct a single state, we can correct an ensemble.

$$\mathcal{N}\left(\left|\tilde{\psi}\right\rangle\left\langle\tilde{\psi}\right|\right) = \sum_i K_i \left|\tilde{\psi}\right\rangle\left\langle\tilde{\psi}\right| K_i^\dagger$$

We will take K_j acting on only one qubit j; but since it is a 2x2 matrix, it can be written as a linear combination of the Pauli basis:

$$K_j \left|\tilde{\psi}_j\right\rangle = \alpha I \left|\tilde{\psi}_j\right\rangle + \beta X_j \left|\tilde{\psi}_j\right\rangle + \gamma Y_j \left|\tilde{\psi}_j\right\rangle + \delta Z_j \left|\tilde{\psi}_j\right\rangle$$

So, any error can be thought of as a superposition of Pauli errors. But syndrome measurement gives a different outcome for all the errors I, X_i, Y_i, Z_i. Doing this measurement destroys the superposition and reduces it to one of the Pauli matrices. But, we know how to deal with an error pattern of one of any of these. So the Shor code works in general.

16 Lecture 16

16.1 General Conditions for Quantum Error Correction (QEC)

We start with the subspace $\mathcal{H}_{code}$, the set of the actual quantum information we want to protect. It is spanned by logical qubits $\{|\tilde{i}\rangle\}$ and we wish to embed this Hilbert space into the physical qubits space $\mathcal{H}_{phys}$. Then we have an alphabet of errors:

$$\Sigma = \{I, E_1, E_2, E_3, \dots\}$$

which we can think of as unitaries in the non-noisy setting or Kraus operators in the quantum channel setting. For the Shor code, the E_i were the 1-qubit Pauli operators. We want to correct any error $E = \sum_i \alpha_i E_i$. When can we correct these errors?

For error correction to be able to happen, we must have for all L, K that if $\langle\tilde{i}|\tilde{j}\rangle = 0$ then $\langle\tilde{i}|E_K^\dagger E_L|\tilde{j}\rangle = 0$. This is a necessary condition, but by itself is not a sufficient condition. Namely take the set of errors where:

$$E_1|\tilde{0}\rangle = E_2|\tilde{0}\rangle, \quad E_1|\tilde{1}\rangle \perp E_2|\tilde{1}\rangle$$

To recover $|\tilde{1}\rangle$ we need to do a syndrome measurement to see whether E_1 or E_2 occurred. But because $E_1|\tilde{0}\rangle = E_2|\tilde{0}\rangle$, intuitively any measurement that does so will affect superpositions $E_i(\alpha|\tilde{0}\rangle + \beta|\tilde{1}\rangle)$.

A stronger condition would be $\langle\tilde{i}|E_K^\dagger E_L|\tilde{j}\rangle = \delta_{ij}\delta_{KL}$. Now, this is sufficient because we can perfectly measure which error E_K actually occurred without affecting the logical state. The projector would be onto the subspace span $\{E_K|\tilde{i}\rangle\}$. Any code that does this is called a **Nondegenerate code**. Note that this condition is not necessary for a code to work, as the Shor code doesn't follow this. For example, $Z_1\left|\tilde{\tilde{0}}\right\rangle = Z_2\left|\tilde{\tilde{0}}\right\rangle$, so the Shor code is degenerate because different errors behave the same way.

We now define the general conditions for error correction:

Theorem 16.1 (Knill-Laflamme Conditions)
Quantum error correction is possible if and only if:

$$\langle\tilde{i}|E_K^\dagger E_L|\tilde{j}\rangle = \delta_{ij} M_{KL}$$

for any matrix M_{KL} that is independent of the indices i, j.

Proof
Only if. Consider a quantum channel $\mathcal{N}(\rho) = \frac{1}{|\Sigma|}\sum_k E_k \rho E_k^\dagger$ which has a uniform distribution over all errors. Recall the Stinespring representation, which wrote $\mathcal{N}(\rho) = \mathrm{tr}_E(V\rho V^\dagger)$ where

$$V|\tilde{i}\rangle = \frac{1}{\sqrt{|\Sigma|}}\sum_k E_k|\tilde{i}\rangle|k\rangle_E$$

Consider $\mathcal{R}$ be the recovery channel, which is the entire process containing both the syndrome measurement and the error-correcting step. We describe it with some Kraus operators $\sum_\mu R_\mu^\dagger R_\mu = I$. We can now give another isometry for the $\mathcal{R}$ channel as V_R with environment E':

$$V_R V|\tilde{i}\rangle = \frac{1}{\sqrt{\Sigma}}\sum_{k,\mu} R_\mu E_k|\tilde{i}\rangle|k\rangle_E|\mu\rangle_{E'}$$

Then the composition of the two channels is:

$$\mathcal{R}\circ\mathcal{N}(\rho) = \sum_{k,\mu} R_\mu E_k \rho E_k^\dagger R_\mu^\dagger$$

If we succeeded, the final state of the system must be $\left|\tilde{i}\right\rangle$.

$$\sum_{k,\mu} R_\mu E_k \left|\tilde{i}\right\rangle |k\rangle_E |\mu\rangle_{E'} = \left|\tilde{i}\right\rangle |\chi\rangle_{E\otimes E'} = \left|\tilde{i}\right\rangle \underbrace{\sum_{k,\mu} \lambda_{k,\mu} |k\rangle |\mu\rangle}_{\text{independent of } \left|\tilde{i}\right\rangle}$$

$$R_\mu E_k \left|\tilde{i}\right\rangle = \lambda_{k,\mu} \left|\tilde{i}\right\rangle$$

Now taking an inner product between $R_\mu E_k \left|\tilde{i}\right\rangle$ and $R_\mu E_k \left|\tilde{j}\right\rangle$, and then summing over the μ's, we have

$$\sum_\mu \left\langle \tilde{j}\right| E_K^\dagger R_\mu^\dagger R_\mu E_L \left|\tilde{i}\right\rangle = \delta_{ij} \sum_\mu \lambda_{K,\mu}^* \lambda_{L,\mu} = \delta_{ij} M_{KL}$$

But the sum can be brought inside and the R_μ are normalized, so the LHS is also $\left\langle \tilde{j}\right| E_K^\dagger E_L \left|\tilde{i}\right\rangle$. This proves that it's necessary for this condition to hold.

If. We will prove sufficiency via explicit construction of R_μ, the recovery operator. Consider the matrix $M_{K,L}$. Let $\hat{e}_\mu = \sum_K e_{\mu,K} \hat{e}_K$, where the $\hat{e}_\mu$'s are the eigenvectors of $M_{K,L}$ with corresponding eigenvalues c_μ. Now, define $M_\mu = \sum_k e_{\mu,k} E_k$. Writing the Knill-Laflamme conditions in terms of M_μ, we have

$$\left\langle \tilde{i}\right| M_\mu^\dagger M_\nu \left|\tilde{j}\right\rangle = \delta_{ij} \delta_{\mu\nu} c_\mu$$

Now, define our recovery operator as follows:

$$R_\mu = \frac{1}{\sqrt{c_\mu}} \sum_i \left|\tilde{i}\right\rangle\left\langle \tilde{i}\right| M_\mu^\dagger$$

Suppose some error happened. Now the state is some superposition of errors E_k, which is also a superposition of M_μ. The state then becomes:

$$\sum_\mu M_\mu \left|\tilde{i}\right\rangle \left|\phi_\mu\right\rangle$$

with no assumptions on the $\left|\phi_\mu\right\rangle$'s. Upon recovery, this becomes

$$\begin{aligned}
\sum_{\nu,\mu} R_\nu M_\mu \left|\tilde{i}\right\rangle \left|\phi_\mu\right\rangle |\nu\rangle &= \sum_{\nu,\mu,j} \frac{1}{\sqrt{c_\nu}} \left|\tilde{j}\right\rangle\left\langle \tilde{j}\right| M_\nu^\dagger M_\mu \left|\tilde{i}\right\rangle \left|\phi_\mu\right\rangle |\nu\rangle \\
&= \sum_{\nu,\mu,j} \delta_{\nu\mu} \delta_{ij} \frac{c_\mu}{\sqrt{c_\nu}} \left|\tilde{j}\right\rangle \left|\phi_\mu\right\rangle |\nu\rangle \\
&= \sum_\mu \sqrt{c_\mu} \left|\tilde{i}\right\rangle \left|\phi_\mu\right\rangle |\nu\rangle
\end{aligned}$$

Now, we just need to show that this is actually a channel. We show that $\sum_\mu R_\mu^\dagger R_\mu = I$ as follows:

$$\begin{aligned}
\sum_\mu R_\mu^\dagger R_\mu &= \sum_{\mu,i,j} \frac{1}{c_\mu} M_\mu \left|\tilde{i}\right\rangle\left\langle \tilde{i}\right| \left|\tilde{j}\right\rangle\left\langle \tilde{j}\right| M_\mu^\dagger \\
&= \sum_{\mu,i} \frac{1}{c_\mu} M_\mu \left|\tilde{i}\right\rangle\left\langle \tilde{i}\right| M_\mu^\dagger \\
&= \text{Proj}\left(\text{span}_{i,\mu} \left\{M_\mu \left|\tilde{i}\right\rangle\right\}\right)
\end{aligned}$$

Finally, we can pad this to the identity by letting $R_0 = I - \sum_{\mu\geq 1} R_\mu^\dagger R_\mu$. Note that this operator does not affect states that we care about, since they live in that subspace spanned by the other R_μ. Thus, the condition is sufficient.

16.2 Stabilizer Codes

Stabilizer codes are a very general way to come up with quantum error-correcting codes. We will first introduce the **Pauli group**.

Definition 16.1 (Pauli group)
The Pauli group on 1 qubit is defined as $\mathcal{G} = \pm\{I, X, iY, Z\}$. This is an 8-element non-Abelian group under matrix multiplication.
The Pauli group on n qubits is defined as $\mathcal{G}_n = \pm\{I, X, iY, Z\}^{\otimes n}$. It contains 2^{2n+1} elements (the plus and minus are applied at the end).

Pauli groups have some nice properties that can be proven simply using casework.

Theorem 16.2 (Properties of Pauli groups)
The Pauli group $\mathcal{G}_n$ on n qubits satisfies the following properties:

1. The Pauli group is indeed a group under matrix multiplication.
2. For $M, N \in \mathcal{G}_n$, either $[M, N] := MN - NM = 0$, or $\{M, N\} := MN + NM = 0$. In other words, they either commute or anti-commute.
3. For $M \in \mathcal{G}_n$, either $M^2 = I$ and $M = M^\dagger$, or $M^2 = -I$ and $M = -M^\dagger$.

Now we are ready to define stabilizer codes. Choose a set $\{M_i\} \subseteq \mathcal{G}_n$. For all i, j, we want $[M_i, M_j] = 0$ and $M_i = M_i^\dagger$. This means the eigenvalues must all be ± 1 and there exists a simultaneous set of eigenstates that diagonalize them. These operators will generate a subgroup $\langle M_i \rangle = S \subset \mathcal{G}_n$ which is Abelian.

For any operator $M \in S$, unless we take $M = I$, half of the eigenvalues of M are 0 and half of them are 1 (they must have trace 0). We then define $\mathcal{H}_{code} = \{|\psi\rangle : \forall i, M_i |\psi\rangle = |\psi\rangle\}$. Note that the dimension of the entire space $\mathcal{H}_{phys}$ is 2^n, so one can show that each condition in $\mathcal{H}_{code}$ halves the dimension. This gives the size of H_{code} to be 2^{n-k}, where $k = |\{M_i\}|$ is the number of generators. Now, via syndrome measurements, we measure all the M_i. Whenever we measure a -1, we have to correct that error, otherwise we do not.

Example 16.1
The Shor code is actually an example of a stabilizer code. The syndrome measurements were

$$\{Z_1Z_2, Z_2Z_3, Z_4Z_5, Z_5Z_6, Z_7Z_8, Z_8Z_9, X_1X_2X_3X_4X_5X_6, X_4X_5X_6X_7X_8X_9\}$$

Note that all of these are members of the Pauli group and they all commute (all even numbers of X and Z).
Exercise: Show that $\left|\tilde{\tilde{0}}\right\rangle, \left|\tilde{\tilde{1}}\right\rangle$ are the only +1 eigenstates of this syndrome measurement.

17 Lecture 17

17.1 More Stabilizer Codes

Recall our definition for a Stabilizer code, succinctly written.

Definition 17.1
A **Stabilizer code** over a space $\mathcal{H}_A$ is a subspace $\mathcal{H}_{code} \subseteq \mathcal{H}_A$ such that

$$\mathcal{H}_{code} = (){\{|\tilde{\psi}\rangle \mid |\tilde{\psi}\rangle \in \mathcal{H}_A,\ \forall i\ M_i |\tilde{\psi}\rangle = |\tilde{\psi}\rangle\}}$$

where $\{M_i\}$ is a set of operators from $\mathcal{G}_n$ that satisfy:

1. They commute. $[M_i, M_j] = M_i M_j - M_j M_i = 0$.
2. They are independent. Namely there is no $M_i = M_{i_1} M_{i_2} \dots$.
3. They are Hermitian. $M_i = M_i^\dagger$

If we have $n-k$ stabilizers, then you have $n-k$ linearly independent conditions, and the dimension of this subspace (the size of the code) is $\dim \mathcal{H}_{code} = 2^{n-(n-k)} = 2^k$. This yields a very succinct way to realize the code as quantum gates.

Theorem 17.1
If we implement $\{M_i\}$ as syndrome measurements and apply appropriate corrections, we can correct any error on up to some number of qubits.
To show this, first consider some error $E_k \in \mathcal{G}_n$. Either $[E_k, M_i] = 0$, which means $M_i E_k |\tilde{\psi}\rangle = E_k M_i |\tilde{\psi}\rangle = E_k |\tilde{\psi}\rangle$, and we have an eigenvalue of +1, or $E_k, M_i = 0$, which means $M_i E_k |\tilde{\psi}\rangle = -E_k M_i |\psi\rangle = -E_k |\psi\rangle$ and we have an eigenvalue of -1.
The Knill-Laflamme conditions say we must show

$$\langle \tilde{i}| E_k^\dagger E_l |\tilde{j}\rangle = \delta_{ij} M_{kl}$$

We study three cases.

1. $E_k^\dagger E_l = M_{i_1} M_{i_2} \dots$, i.e. it is in the group generated by the M_i's. Then since $|\tilde{i}\rangle$ are eigenvectors of the M_i's with eigenvalue +1, they also have the same eigenvalue for $E_k^\dagger E_l$. So $\langle \tilde{i}| E_k^\dagger E_l |\tilde{j}\rangle = |\tilde{i}\rangle\langle \tilde{j}| = \delta_{ij}$.
2. $\{E_k^\dagger E_l, M_i\} = 0$ for some i. Then $E_k^\dagger E_l |\tilde{j}\rangle$ is an eigenvector of M_i with eigenvalue -1 (by the above), but $|\tilde{i}\rangle$ is an eigenvector with eigenvalue +1, so they must be orthogonal, so $|\tilde{i}\rangle E_k^\dagger E_l \tilde{j} = 0 = 0\delta_{ij}$.
3. $[E_k^\dagger E_l, M_i] = 0$ for all i, but is not in case 1. Then $E_k^\dagger E_l |\tilde{j}\rangle \in \mathcal{H}_{code} \neq |\tilde{j}\rangle$ in general. So we cannot prove the Knill-Laflame conditions.

In the first and second case, we are good. Suppose we find these $\{M_i\}$ such that for all i, $[M, M_i] = 0$ means that M acts on at least d qubits, which we will call the **distance** of the code. We thus limit our discussion to $E_k = \{I, \text{Paulis on } (d-1)/2 \text{ qubits}\}$. Then $E_k^\dagger E_l$ acts on at most $d-1$ qubits, meaning $\mathcal{H}_{code}$ satisfies Knill-Laflamme conditions.

Example 17.1
To see that d qubits in action, let's take a look at the Shor code.

$$\{M_i\} = \{Z_1Z_2, Z_2Z_3, Z_4Z_5, Z_5Z_6, Z_7Z_8, Z_8Z_9, X_1X_2X_3X_4X_5X_6, X_4X_5X_6X_7X_8X_9\}$$

To commute with all the stabilizers, it would have to commute

Definition 17.2
A quantum error-correcting code with n physical qubits, k logical qubits and distance d is denoted as a $[\![n, k, d]\!]$ quantum error-correcting code.

So, we have shown the Shor code is a $[\![9, 1, 3]\!]$ quantum error-correcting code. Here are a few:

Example 17.2
The **Steane code** is a $[\![7, 1, 3]\!]$ stabilizer code with stabilizers:

$$\{Z_1Z_3Z_5Z_7, Z_2Z_3Z_6Z_7, Z_4Z_5Z_6Z_7, X_1X_3X_5X_7, X_2X_3X_6X_7, X_4X_5X_6X_7\}$$

Clearly the Z_i's commute, but to see the Z's and X's commute, notice that each product of Z's shares an even number of factors with each product of X's.
The **Five-qubit code** is a $[\![5, 1, 3]\!]$ stabilizer code with stabilizers:

$$\{X_1Z_2Z_3X_4, X_2Z_3Z_4X_5, X_1X_3Z_4Z_5\}$$

One may ask if we can do better. The answer is no.

Theorem 17.2
There is no $[\![4, 1, 3]\!]$ code.

Proof
Consider the set of errors E_k that only act on the first two qubits, so $E_k^\dagger E_l$ acts on first two qubits. Consider an arbitrary state $|\tilde{\psi}\rangle$. Send the first two physical qubits to Alice and the last two physical qubits to Bob. Then Both Alice and bob could recover $|\tilde{\psi}\rangle$, but this would clone the state! So such a code could not exist.

17.2 Toric Code

Consider a square lattice. We are going to place one physical qubit on each edge of the lattice. Further we add a toroidal topology, so the top of the lattice is connected to the bottom and the right of the lattice is connected to the left. Let v be a vertex, e be an edge (physical qubit) and p be a face (plaquette) of this lattice. Define $e(p) = \{e : e = \{p, \cdot\}\}$ and $A_v = \prod_{e(p)} X_e$. Since each vertex has 4 incident edges, there are 4 X operators and $B_p = \prod_{e(p)} Z_p$ where we mean incident edges in the dual graph, i.e. the edges surrounding the face. Again, there are 4 per face, so this is very explicit. Check the commutation, since X's and Z's commute with themselves:

$$[A_v, A_{v'}] = 0, [B_p, B_{p'}] = 0$$

Now, we explore $[A_v, B_p]$. They clearly commute if they share no edges. If they share an edge, then v must be a corner of p, so they share exactly two edges. But each of these introduce a minus sign when commuted, which cancel to 0, i.e. $[X_1X_2, Z_1Z_2] = 0$. Thus, all these operators commute and form a valid set of stabilizers. Suppose the lattice has size L. What is the number of logical qubits here?

$$\dim \mathcal{H}_{phys} = 2^{n_e} = 2^{2L^2}$$

How many stabilizers do we have? We have L^2 vertices and L^2 plaquettes, giving $2L^2$ stabilizers? But wait, not all of these are independent, because $\prod_v A_v = I$ (by handshaking, every vertex contributes two copies of every X operator, and $X_i^2 = I_i$, so this leads to the identity). Similarly, $\prod_p B_p = I$. Thus, we only have $2L^2 - 2$ independent stabilizers. So $k = 2$ logical qubits and $\dim \mathcal{H}_{code} = 2^2 = 4$. Let's find its distance. Suppose $[E_k^\dagger E_l, A_v] = 0$ for all v and

$[E_k^\dagger E_l, B_p] = 0$ for all p. We know that for some set e_X, e_Z:

$$E_k^\dagger E_l = \pm \prod_{i \in e_X} X_i \prod_{j \in e_Z} Z_j$$

So $[E_k^\dagger E_l, A_v] = 0$ if and only if e_Z and $e(v)$ share $0, 2$, or 4 edges. Such a graph must be a set of closed loops. Similarly, $[E_k^\dagger E_l, B_p] = 0$ if and only if e_X and $e(p)$ share an even number of edges. This means the dual graph must be a set of closed loops.

18 Lecture 18

18.1 Toric Code, Continued

Recall that we had a toroidal lattice (the top connects to the bottom and left connects to the right) on which, we placed a qubit on each edge e. These gave rise to vertex and plaquette operators, $A_v = \prod_{e \in e(v)} X_e$ and $B_p = \prod_{e \in e(p)} Z_e$. We saw these stabilizers gave a $[\![n = 2L^2, k = 2, d = ??]\!]$ code. Let's find that last parameter.

Theorem 18.1
The Toric code is a $[\![2L^2, 2, L]\!]$ code.

Proof
We need to find the minimum number of qubits $E_k^\dagger E_l$ acts on such that $[E_k^\dagger E_l, A_v] = [E_k^\dagger E_l, B_p] = 0$ but is not generated, i.e. $E_k^\dagger E_l \neq \prod_{v_i} A_{v_i} \prod_{p_j} B_{p_j}$. We saw that $E_k^\dagger E_l = \prod_{e \in e_x} X_e \prod_{e \in e_z} Z_e$ commuting with all stabilizers was equivalent to e_z forms a set of closed loops and and e_x forms a set of closed loops in the dual lattice.
But this is not a hard condition, consider $E_k^\dagger E_l = B_p$ (i.e. a closed loop of a square). But this doesn't tell us anything about the distance, because it's still generated by our set of stabilizers. So, we must ask, what is the smallest operator $E_k^\dagger E_l$ which is not generated. The e_z and e_x are independent, so the smallest operator has only Z's or only X's. Let's consider only Z's without loss of generality. A loop that cannot we written as a product of vertex and plaquette operators is called a **noncontractible loop**. Such a loop would be a vertical or horizontal line, remember that the edges are glued together. We can also form a noncontractible loop using X's by using the dual lattice (take the horizontal edges along any vertical line that is not grid-aligned). So we can correct up to L distance.

Now let's consider the following commutators:

$$\left[\prod_{e \in \text{vertical loop}} X_e, \prod_{e \in \text{vertical loop}} Z_e \right] = 0$$

$$\left\{ \prod_{e \in \text{horizontal loop}} X_e, \prod_{e \in \text{vertical loop}} Z_e \right\} = 0$$

By these commutation relations, we can identify logical operators as the following:

$$\tilde{X}_1 = \prod_{e \in \text{v. loop}} X_e$$

$$\tilde{X}_2 = \prod_{e \in \text{h. loop}} X_e$$

$$\tilde{Z}_1 = \prod_{e \in \text{h. loop}} Z_e$$

$$\tilde{Z}_2 = \prod_{e \in \text{v. loop}} Z_e$$

Such an error-correcting code is termed as using **Topological Quantum Error Correction**.

18.2 Fault Tolerance

We want to build a quantum computer with quantum error correction. But the error-correction itself is made up of quantum gates, and may itself be faulty. So we must somehow resolve this. The first results we care about are threshold theorems. There are many for different error-correcting codes, but they all are statements of the following form:

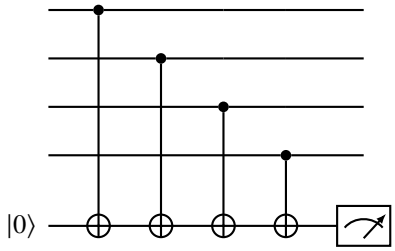
PseudoTheorem 18.1 (Threshold Theorem)
Call p_T the thereshold probability, which depends on the code we use and our fault-tolerance procedure. If you can build a physical quantum computer with "error probability per gate" $p < p_T$ This means we can build a logical quantum computer with error probability per gate $p' < p$.

As a consequence, we can iteratively build more and more fault-tolerant quantum computers and make the error rate arbitrarily small. The first proofs had threshold $p_T = O(10^{-7})$, which was useless for practice. Now we have gotten $p_T = O(10^{-3})$.

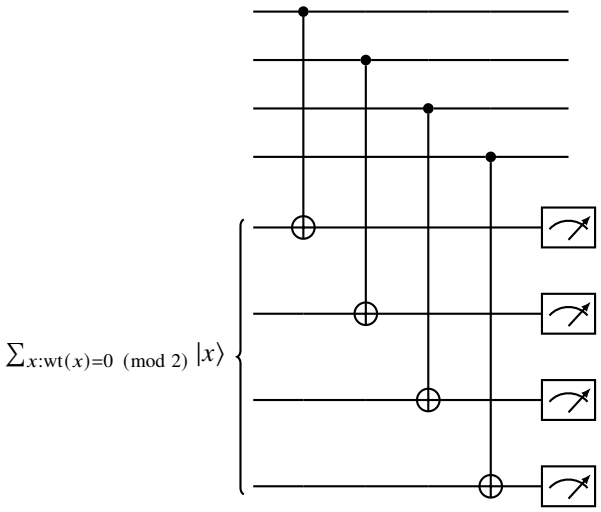
So how does this model work? We have two main sources of issues:

1. Carrying out error-correction can create new errors.
2. Implementing logical gates can create errors.

On issue 1, a syndrome measurement can not only create new errors, but can spread errors. Suppose we'd like to measure $Z_1Z_2Z_3Z_4$ on some four physical qubits. The usual way to do this is:

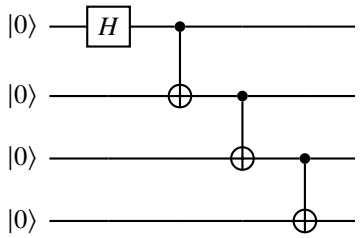


But suppose there's an error on the Ancilla qubit. Then this single error "infects" all four physical qubits. We want an approach to do these measurements so any single-qubit error cannot affect more than 1 physical qubit. Let's instead use 4 physical qubits:

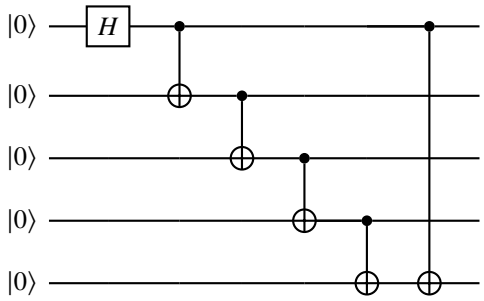


Now after measurement, if there are an even number of 1s then there is no error and if there are an odd number, there is an error, but we learn no other information about the physical qubits. But now there's another issue, how do we create this superposition over $|x\rangle$ in a fault-tolerant way? We claim

$$\sum_{x:\mathrm{wt}(x)=0 \pmod 2} |x\rangle = H^{\otimes 4} \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)$$



This does not work if we have an error on the first or fourth qubit. Let's add another ancilla qubit to check for this:



If we measure 0, then no dangerous errors happened, so just keep trying until we measure 0.

Now on issue 2, the stupid way would be to decode logical qubit into physical qubit, implement the gate, then reencode. But during the gate, we could get an error and be unable to correct it. Thus, we need to implement the logical gates while everything is encoded, like we saw in the toric code $\tilde{X}_1 = \prod_{e \in \text{horiz. loop}} X_e$. So, there is no error spreading because each gate is a 1-qubit gate. This is called a **transversal logical gate** and so only acts as a product over physical qubits.

Definition 18.1
A gate U is said to be a **Clifford gate** if the group of Paulis is invariant under its conjugation.

$$U\mathcal{G}_n U^\dagger = \mathcal{G}_n$$

Theorem 18.2
For any stabilizer code, any Clifford gates U can be implemented transversally.

Theorem 18.3
The set of all Clifford gates can be generated by:

$$\langle H, CNOT, S \rangle$$

where $S = |0\rangle\langle 0| + i\,|1\rangle\langle 1|$.

Theorem 18.4
Clifford gates can be simulated efficiently using classical computers.

If all we can do is Clifford gates, what was the point?

Theorem 18.5 (Eastin-Knill)
There is no quantum-error correction where a universal gate set can be implemented transversally.

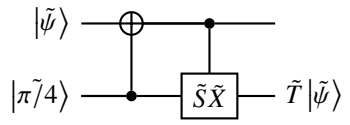
But actually, for any stabilizer code, we only need to implement one non-Clifford gate to do any computation we want.

$$T = |0\rangle\langle 0| + e^{i\pi/4}\,|1\rangle\langle 1|$$

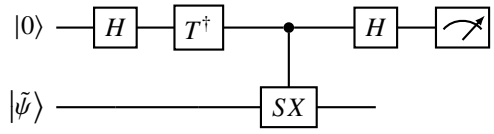
Now to implement any other gate we first consider the magic state:

$$\left|\tilde{\pi/4}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|\tilde{0}\right\rangle + e^{i\pi/4}\left|\tilde{1}\right\rangle\right)$$

Then this is the implementation of that gate:



But note that $\left|\tilde{\pi/4}\right\rangle$ is an eigenstate of SX which is Clifford.



However, this is not fault-tolerant, because error on the first ancilla qubit propogates to all of the logical qubits. This can be fixed using exactly the same techniques we used for the syndrome measurement.