CSE439 Week 11: Grover's Algorithm (ch. 13), then into Ch. 14

All of our previous quantum algorithms have been ones where an n-qubit Hadamard transform has been applied once, then an oracle gate or other computation to create a functional superposition

$$\sum_{x \in \{0,1\}^n} |x\rangle |f(x)\rangle,$$

and then *one* further transform---Hadamard or Fourier---before measuring the entire output. [Footnote: the notation $|xf(x)\rangle$ for the body of the sum is equivalent.] Further iterations are managed by a routine with *classical* control. Grover's algorithm, however, has successive quantum stages that each use two banks of Hadamard gates. The $2^n \times 2^n$ matrices $\mathbf{H}^{\otimes n}$ are just as easy as any other in a "Schrödinger-style" simulation where you multiply matrices. But in a "Feynman-style" simulation where we count nondeterministic witness strings, the repeated Hadamard transforms mushroom the witness space. (This is why the `groverDemo` in my simulator has not been implemented yet.)

Grover's algorithm as originally presented applies only at "witness scale": a space of $N = 2^q$ potential witness strings using $q = q(n)$ qubits, **not** N separate physical locations as commonly talked about. Whether it can apply to N physical sites with $\widetilde{O}\left(\sqrt{N}\right)$ **effort** is IMHO controversial. However, at witness scale, there aren't even $\sqrt{N}$ physical sites, only q qubits with basis vectors $|0^q\rangle$ through $|1^q\rangle$. A solution set $S \subseteq \{0,1\}^q$ is represented by the "hit vector" $\mathbf{h_S}$ defined by

$$\mathbf{h}_S(y) = \begin{cases} \dfrac{1}{\sqrt{|S|}} & \text{if } y \in S \\ 0 & \text{otherwise} \end{cases}.$$

This is just the normalized sum of the basis vectors corresponding to strings in S. *Except:* if S is empty, then this would be the zero vector in $\mathbb{C}^N$, which is not a legal quantum state. There is a further worm in this apple:

- There are $2^N = 2^{2^q}$ different possible subsets S.
- Thus it seems that each hit vector $\mathbf{h}_S$ carries N bits of information.
- However, we are using only $q \ll N$ qubits, and we need to remind ourselves about:

Holevo's Theorem: It is not possible to extract more than q bits of classical information from any q-qubit quantum state.

Thus, like the situation with graph states, the quantum representation of solution sets is inevitably ***lossy***.

This is part reason for Lov Grover's original attention only to singleton sets $S = \{y\}$, whereupon we simply have $\mathbf{h_S} = |y\rangle$. Then distinguishing among the 2^q possibilities (all of them not the empty set) involves only q bits of information. Any setting that allows $|S| > 1$ involves some information smearing. The final point here is that the measurement at the end of the algorithm will give you just *one* witness, not necessarily the whole set of them. When $S = \{y\}$ it is the whole set, but otherwise not.

At witness scale, the running time is not sub-linear but merely quadratically sub-exponential: $\widetilde{O}\left(\sqrt{N}\right) = n^{O(1)}2^{q(n)/2}$, which is still 2-to-the-linear exponential time, not even $2^{q(n)^{1/2}}$. Here the multipler---which is the time per iteration---includes the polynomial gate count $s = s(n) = n^{O(1)}$. As an aside, I am skeptical that this is a true measure of quantum effort. Well, we should examine the quantum circuits, after seeing the idea of the algorithm.

How Grover Search Works

Grover's algorithm actually operates completely within a 2-dimensional subspace of $\mathbb{C}^N$. The subspace is spanned by two vectors: $\mathbf{h_S}$ and the vector $\mathbf{j} = \mathbf{H}^{\otimes q}|0^q\rangle$. (Unless $S = \{0,1\}^q$ *in toto*, which makes them equal.) We do not know $\mathbf{h_S}$ in advance, but we do know $\mathbf{j}$. The "miss" vector $\mathbf{m}_S = \mathbf{h}_{\sim S}$ also belongs to the subspace, since it equals

$$\frac{\sqrt{N}\cdot\mathbf{j} - \sqrt{|S|}\cdot\mathbf{h_S}}{\sqrt{N-|S|}}, \quad \text{so that} \quad \mathbf{j} = \frac{\sqrt{N-|S|}}{\sqrt{N}}\mathbf{m}_S + \frac{\sqrt{|S|}}{\sqrt{N}}\mathbf{h}_S.$$

We don't know $\mathbf{m}_S$ either, but provided S is given by a polynomial-time decidable witness predicate $R(x,y)$ of our problem instance x, then we can reflect around it by means of the **Grover oracle**

$$U_R[xy,xy] = (-1)^{R(x,y)} = \begin{cases} -1 & \text{if } R(x,y) \\ 1 & \text{if } \neg R(x,y) \end{cases}.$$

When x is fixed, the Grover oracle drops down to an $N \times N$ diagonal matrix G_x with entry $G_x[y,y] = -1$ if $y \in S$ and $G_x[y,y] = 1$ otherwise. To compute it, we can apply an idea that the textbook calls "flipping a switch" in section 6.5 but might be better called the idea of using an extra qubit as a *catalyst*. The catalyst is that we initialize the extra qubit not to $|0\rangle$ or $|1\rangle$ but to

$$\mathbf{d} = \mathbf{H}|1\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right).$$

We can create a quantum circuit C_0 of deterministic gates only (Toffoli plus constant initializations) for the reversable form of the Boolean function $f_x(y) = R(x,y)$, which is the $(q+1)$-bit function $F_x(yb) = y(b \oplus f_x(y))$. Now define $g_x(y) = C_0\left(|y\rangle \otimes \mathbf{d}\right)$ using our catalyst. We get

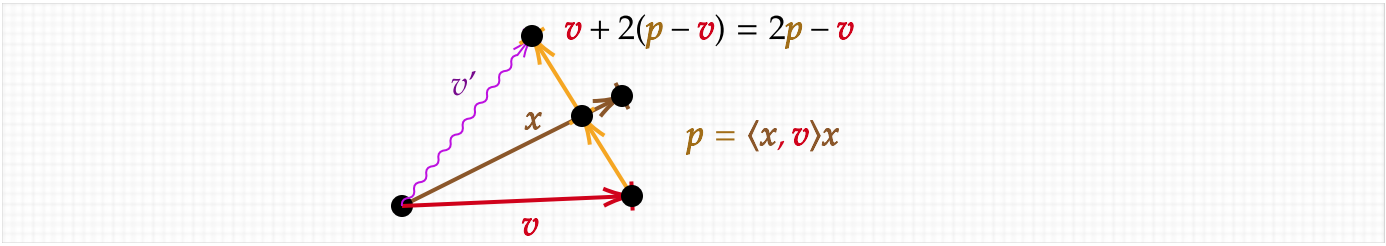
$$g_x(y) = C_0(|y\rangle\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) = \frac{C_0|y0\rangle - C_0|y1\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}}\left(|y\rangle|f_x(y)\rangle - |y\rangle|\neg f_x(y)\rangle\right)$$

$$= \begin{cases} \dfrac{|y1\rangle - |y0\rangle}{\sqrt{2}} & \text{if } f_x(y) = 1 \\ \dfrac{|y0\rangle - |y1\rangle}{\sqrt{2}} & \text{if } f_x(y) = 0 \end{cases} = \begin{cases} |y\rangle \otimes (-\mathbf{d}) & \text{if } R(x,y) \\ |y\rangle \otimes \mathbf{d} & \text{if } \neg R(x,y) \end{cases} = (-1)^{R(x,y)}|y\rangle \otimes \mathbf{d}$$

If we "throw away" the last qubit (say by measuring it and ignoring the result) then we get the Grover oracle action on the first q qubits. So for polynomial-time witness predicates $R(x, y)$, the Grover oracle is feasible to compute.

The key next point is that in the geometry of the 2-dimensional space, the Grover oracle represents reflection around the *miss* vector $\mathbf{m}_S$. Note first that $G_x\mathbf{m}_S = \mathbf{m}_S$ because no nonzero entry gets negated. And $G_x\mathbf{h}_S = -\mathbf{h}_S$ because all the nonzero entries get negated. Therefore the action of G_x in this space is *reflection about* $\mathbf{m}_S$.

The other operation we want is reflection about $\mathbf{j}$. In general, reflection of a vector v around a vector x involves first taking the projection of v onto x, which is $\langle v, x\rangle x$. Then we want to move v by twice the difference of that to v:



The matrix operator that creates the projection of an argument v along x is the **outer product** $|x\rangle\langle x|$, whose $[i, j]$ entry is $x_i\overline{x_j}$. The Dirac notation is especially handy here, because we can do

$$|x\rangle\langle x| \cdot |v\rangle = |x\rangle\langle x|v\rangle = \langle x, v\rangle|x\rangle.$$

So the operator that creates the reflection is $2|x\rangle\langle x| - \mathbf{I}$. In the case $x = \mathbf{j}$ this is given by the matrix $2\mathbf{J} - \mathbf{I}$ where each entry of $\mathbf{J}$ is $\frac{1}{N}$ and $\mathbf{I}$ is the $N \times N$ identity matrix.

Because we are talking about exponential-sized matrices, it is relevant to ask about the feasibility of computing their actions. An equation by which to build the reflection about $\mathbf{j}$ is

$$2\mathbf{J} - \mathbf{I} \;=\; \mathbf{H}^{\otimes q}(-1)^{NOR(1..q)}\mathbf{H}^{\otimes q}.$$

The $(-1)^{NOR(1..q)}$ is implemented via a controlled-$\mathbf{Z}$ gate on one qubit with controls on the other $(q-1)$ qubits---it doesn't matter which, as the gate is symmetric. By itself, that gate computes $(-1)^{AND(1..q)}$, so it is sandwiched between two banks of $\mathbf{NOT}$ gates to get the action of NOR. To see why this works, consider first that on any basis input $|x\rangle$, $\mathbf{H}^{\otimes q}|x\rangle = \frac{1}{\sqrt{N}}\sum_y (-1)^{x\odot y}|y\rangle$. Applying the $(-1)^{NOR(1..q)}$ gives

$$\frac{1}{\sqrt{N}}\sum\nolimits_{y\neq 0^q}(-1)^{x\odot y}|x\rangle \;+\; \frac{(-1)}{\sqrt{N}}(-1)^{x\odot 0^q}|0^q\rangle \;=\frac{1}{\sqrt{N}}\sum\nolimits_{y}(-1)^{x\odot y}|y\rangle \;-\; \frac{2}{\sqrt{N}}|0^q\rangle$$

Applying $\mathbf{H}^{\otimes q}$ again gives

$$\frac{1}{N}\sum\nolimits_y\sum\nolimits_z(-1)^{x\odot y}(-1)^{z\odot y}|z\rangle \;-\; \frac{2}{N}\sum\nolimits_z(-1)^{z\odot 0^q}|z\rangle \;=\frac{1}{N}\sum\nolimits_y\sum\nolimits_z(-1)^{(x\oplus z)\odot y}|z\rangle \;-\; \frac{2}{N}\sum\nolimits_z|z\rangle$$

Now the outer sum over y in the first term vanishes except when $z = x$, so we get

$$\frac{1}{N}\sum\nolimits_y|x\rangle \;-\; \frac{2}{N}\sum\nolimits_z|z\rangle \;=\; |x\rangle \;-\; \frac{2}{N}\sum\nolimits_z|z\rangle \;=\; (\mathbf{I} \;-\; 2\mathbf{J})|x\rangle.$$

This is (-1) times what we expected, but the global scalar does not matter. The last thing to say is that whenever v belongs to our 2-dimensional subspace, the reflection of v around $\mathbf{j}$ stays within it.

[Thursday's lecture will pick up here, revisiting the diagram but this time to emphasize what we do not know and why we cannot take shortcuts.]

The Search Process

Let α stand for the angle between $\mathbf{j}$ and $\mathbf{m}_S$. Then $\alpha = \cos^{-1}\langle \mathbf{j}, \mathbf{m}_S\rangle = \sin^{-1}\langle \mathbf{j}, \mathbf{h}_S\rangle$. When $|S| = o(N)$, then α is small enough to use the estimate $\sin\alpha \approx \alpha$, so we get

$$\alpha \;=\; \sin^{-1}\langle \mathbf{j}, \mathbf{h}_S\rangle \;\approx\; \langle \mathbf{j}, \mathbf{h}_S\rangle \;=$$

$$\sum_y [y \text{ is a solution}]\cdot\frac{1}{\sqrt{N}\sqrt{S}} \;=\; S\cdot\frac{1}{\sqrt{N}\sqrt{S}} \;=\; \frac{\sqrt{|S|}}{\sqrt{N}} \;=\; \sqrt{\frac{|S|}{N}}.$$

The number of iterations (each a pair of reflections) we will need is about $\frac{\pi/2}{2\alpha} = \frac{\pi}{4\alpha} \approx \frac{\pi}{4}\sqrt{\frac{N}{S}}$. This is always about the square root of the expected time for guessing uniformly at random and verifying. If

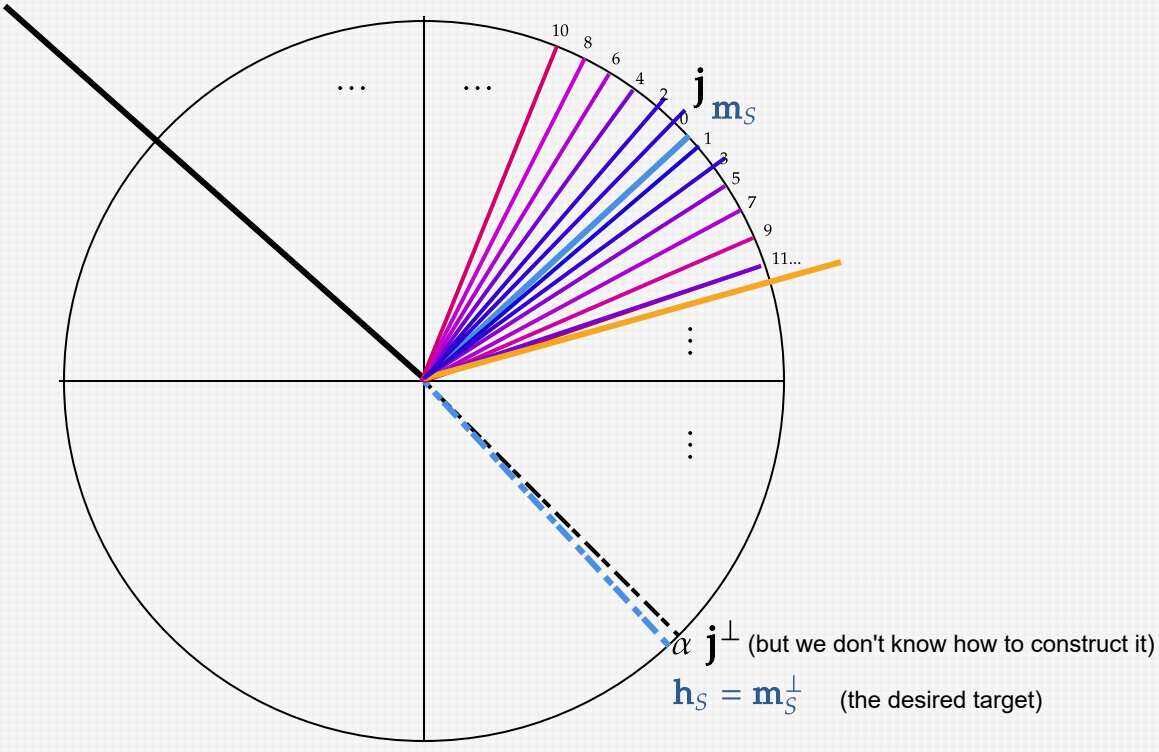
we know $|S|$, then we know how many iterations to make before measuring; if we don't know $|S|$, then there are further tradeoffs discussed later. In any event, unless $|S| = \Omega(N)$, we have $\alpha = o(1)$, so that the angle α is best pictured as very small. When $|S| \leq \sqrt{N}$, we have

$$\frac{1}{\sqrt{N}} \leq \alpha \leq \frac{1}{\sqrt[4]{N}}$$

as the most relevant range of angles. Now to summarize what we know and don't know:

1. We know a vector $\mathbf{j}$ in the two-dimensional subspace H generated by the hit vector $\mathbf{h}_S$ and its orthogonal complement, the miss vector $\mathbf{m}_S$.
2. The goal is to build a quantum state ϕ whose vector is within ϵ of $\mathbf{h}_S$, so that measuring ϕ will with probability $\approx 1 - \epsilon$ yield a member of S.
3. We know that $\mathbf{j}$ is close to $\mathbf{m}_S$, so that $\mathbf{j}^\perp$ is close to $\mathbf{h}_S$ (or opposite to $\mathbf{h}_S$---either way, measuring $\mathbf{j}^\perp$ would yield a solution whp.), but we have no idea how to construct $\mathbf{j}^\perp$ within H.
4. What we do have are feasible circuit components computing reflection around $\mathbf{m}_S$ and reflection aound $\mathbf{j}$ that stay within H.
5. If we know $|S|$, then we know the number of iterations that produces a vector ϕ closest to $\mathbf{h}_S$. Moreover, ϕ will be within angle α of $\mathbf{h}_S$.

Here is a diagram of the iteration process. It is different from most other diagrams by emphasizing the smallness of α and not giving the impression that $\mathbf{j}^\perp$ is knowable by aligning it with vertical or horizontal axes. The iteration starts by reflecting the known vector $\mathbf{j}$ around $\mathbf{m}_S$. The next five iterations (each a rotation by 2α effected by two reflections) are shown and color-coded.



It may seem strange that we cannot jump straight to $\mathbf{j}^{\perp}$ from $\mathbf{j}$ or otherwise leverage the initial proximity to $\mathbf{m}_S$ in a way that would at least allow taking bigger steps toward $\mathbf{h}_S$ than repeated rotation by 2α. It looks even more enticing upon realizing that getting within 45° of $\mathbf{h}_S$, that means anywhere in the lower-right quadrant shown, gives at least a $\sin^2\left(\frac{\pi}{4}\right) = \frac{1}{2}$ chance of the measurement giving a string in S.

The picture makes it look like we could hit that quadrant quickly just by throwing darts at it. But the point is that the "dartboard" H is hidden inside a vastly higher dimensional space, and we have no direct information besides the $\mathbf{j}$ vector of how it lies. In fact, the above process is tightly optimal.

The Case of General Solution Count

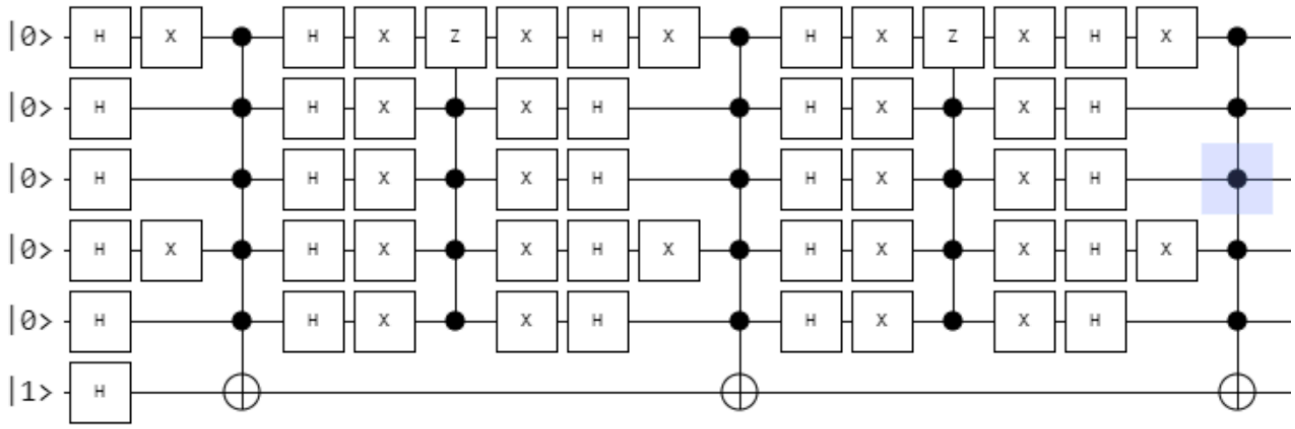
If $|S|$ is unknown, we can guess a stopping time $t \leq \sqrt{N}$ uniformly at random. Now the "dartboard" reasoning works in our favor since everything happens within the subspace H, and the expected time to find a solution is only a constant factor greater than when $|S|$ is known. Namely:

- Do t pairs of reflection steps around $\mathbf{j}$ and around $\mathbf{m}_S$.
- If t happens to be $\gg |S|$ this may gyrate multiple times around the circle. But:
- Except when S is everything, this gives about at least a 0.25 chance of ending in the quadrant of within 45° of $\mathbf{h}_S$.
- If the trial ends up in the quadrant, it has at least a 0.5 chance of getting a true solution y from the measurement. The y can be verified by computing $R(x, y)$ deterministically.
- Thus each random trial over t has at least a 0.125 chance of ultimate success.

This argument is sloppier than it needs to be, but it's enough to conclude that a solution can be expected within a constant number of trials, regardless of the size of S. Thus the expected time to find a solution remains $O\left(\sqrt{N}\right)$.

Circuit Implementation and Problematic Aspects

The Grover oracle is deterministic except for the single Hadamard gate used to initialize the catalyst qubit to the difference state $\mathbf{d}$. We do not have to re-initialize it, however, because the output after the evaluation remains $(-1)^{R(x,y)}|y\rangle \otimes \mathbf{d}$. The issue is the reflection about $\mathbf{m}_S$. Done straightforwardly, it is heavy on the $\mathbf{H}$ gates, as evinced by the following example which can be created in Davy Wybiral's quantum web applet:



(show all)

0.11132812+0.00000000i	|011000>	1.2394%
-0.11132812+0.00000000i	|011001>	1.2394%
0.33007813+0.00000000i	|011010>	10.8952%
-0.33007813+0.00000000i	|011011>	10.8952%
0.11132813+0.00000000i	|011100>	1.2394%

Here the Grover oracle is $\bar{x}_1 \wedge x_2 \wedge x_3 \wedge \bar{x}_4 \wedge x_5$ giving $S = \{01101\}$. This is implemented as a multi-controlled flip of the catalyst line (where a single $\mathbf{H}$ follows the ancilla $|1\rangle$ value) with $\mathbf{X}$ gates to make $\bar{x}_1$ and $\bar{x}_4$. The initial bank of Hadamards on the first five qubits is to create the $\mathbf{j}$ vector on them. The four other banks, however, are for the two reflections about $\mathbf{j}$. The angle α is $\sin^{-1}\left(1/\sqrt{32}\right) = 0.1777...$ radians. The desired number of iterations is $\frac{\pi}{4\alpha} = 4.42$; the diagram counts as 2.5 iterations. This is close enough to show more probability accumulating on the string 01101 on the first five qubits.

If we make a brute-force algebraic or logical simulation out of this, however, the Hadamard gates for the reflections give rise to 20 new variables. The number of Feynman paths grows by a factor of more than 1,000 per iteration. (This also causes major branching in the witness space for problem 3 on assignment 4.) This growth would quickly choke the path-counting simulation written in C++ which I've demo'ed. [Hence `groverDemo` is not yet coded. The hope is to perform logical simplifications of the representation of the current quantum state so as to combine paths and reduce the branching factor, but results so far have not been promising.]

The multi-controlled Z gate has its own element of excess. Yes, OK, the Grover oracle in this case is also multi-controlled, but one expects to expend more effort on it---and it could be a larger network of gates with only one control each. The reflection about $\mathbf{j}$, however, really uses all the controls. IBM researchers have found even the double-controlled Toffoli gate to be difficult to engineer, which is why their preferred basis consists of $\mathbf{H}$, $\mathbf{CNOT}$, and the $\mathbf{T}$ gate.

Chapter 14: Qubit Representations, Physical States, and Operators

A **qubit** is a physical system whose **state** ϕ is described by a pair (a, b) of complex numbers such that $|a|^2 + |b|^2 = 1$. The components of the pair *index* the *basic outcomes* 0 and 1. There are two ways we can gain knowledge about the values a and b:

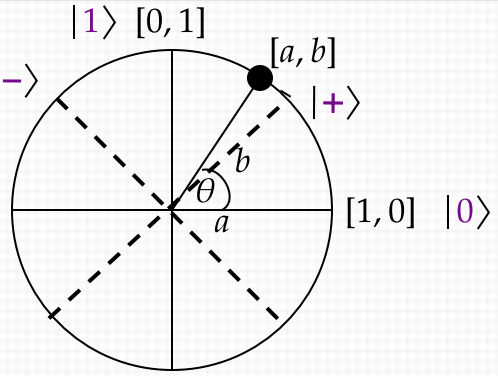
- We can **prepare** the state from the known initial state $e_0 = (1, 0)$ by known quantum operations, which here can be represented by 2×2 matrices.
- We can **measure** the state (with respect to these basic outcomes), in which case:
  - We either **observe** 0, whereupon the state becomes e_0, or we observe 1, in which case the state becomes $e_1 = (0, 1)$.
  - The probability of observing 0 is $|a|^2$, of getting 1 is $|b|^2$. Called the **Born Rule**, for Max Born.

If both a and b are real numbers, then we can picture the qubit as a point on the unit circle in $\mathbb{R}^2$:

$|a|^2 + |b|^2 = 1$

$a = \cos\theta$

$b = \sin\theta$



(transpose notation omitted here)

The qubit state $\phi = [a, b]$ represents $ae_0 + be_1 = a[1, 0] + b[0, 1] = a|0\rangle + b|1\rangle$

If $\theta = \frac{\pi}{3}$ then $\cos\theta = \frac{1}{2}$, so $|a|^2 = 0.25$. And $\sin\theta = b = \frac{\sqrt{3}}{2}$ so $|b|^2 = 0.75$. Note that $a = \langle\phi|0\rangle$ and $b = \langle\phi|1\rangle$. What the measurement does is **project** onto the standard basis.

We can get different probabilities by projecting onto a different basis. Note that

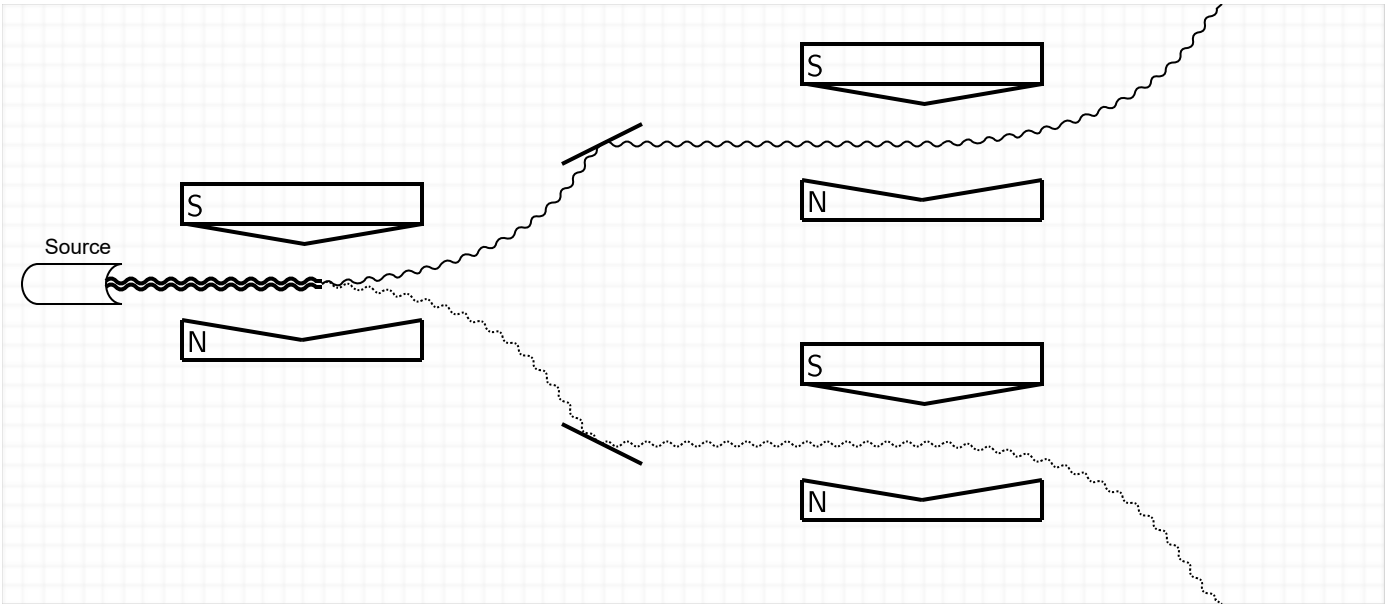
$$\langle\phi|+\rangle = \frac{1}{\sqrt{2}}\left(\frac{1}{2}\cdot 1 + \frac{\sqrt{3}}{2}\cdot 1\right) = \frac{1+\sqrt{3}}{2\sqrt{2}} = \frac{2.732...}{2.828...} = 0.9659...$$

and squaring that gives just over 0.933. Thus, this particular quantum state ϕ gives a higher probability of one result when measured in the $|+\rangle, |-\rangle$ basis---and a near-zero probability of the other result.

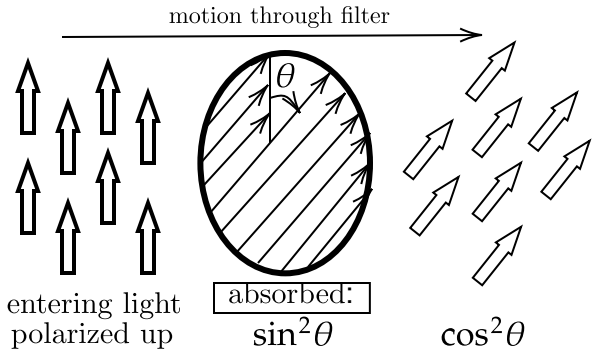
What *happens* to ϕ after a measurement? The full picture is much debated, but the local happening is clear: ϕ becomes the basis state corresponding to the result obtained. The fact that we---humans---can *elect* to **measure in** a particular choice of basis will be a major component of quantum communication protocols and the **CHSH Game** on-tap later in Chapter 14. The "election" part is as

easy as twirling a polaroid filter (if that is *free will*, mind you).

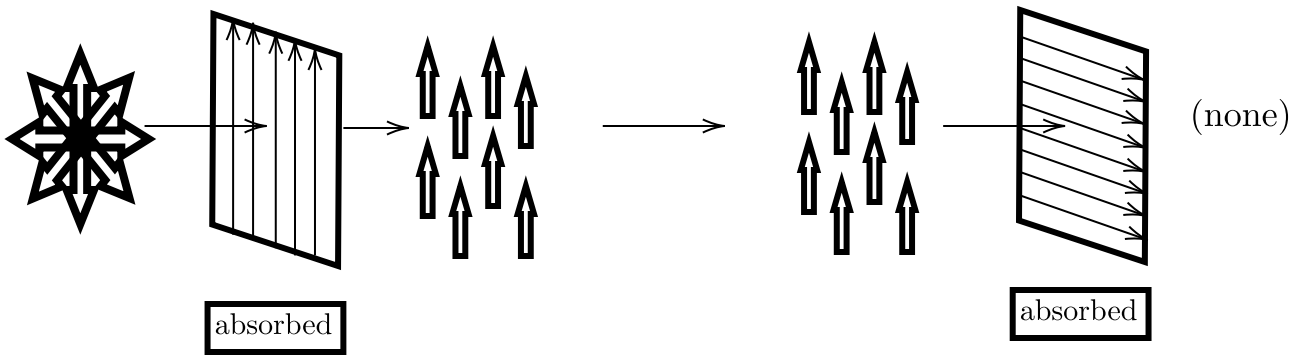
(1) That the particles' states become basis states in the particular measurement frame is shown by the **Stern-Gerlach experiment**. In the setup, the measurable physical state "spin up" is denoted by $|\uparrow\rangle$ and can be treated like $|0\rangle$. There is a distinct physical state called "spin down" and denoted by $|\downarrow\rangle$, which plays the role of $|1\rangle$. These are the only two distinguishable outcomes that manifest when a magnetic field acts on the particle (relative to the orientation of the field; incidentally, "spin" is not-rotation per-se). Once a particle "chooses" between $|\uparrow\rangle$ or $|\downarrow\rangle$, that is its state upon going through a second Stern-Gerlach device with the same orientation.

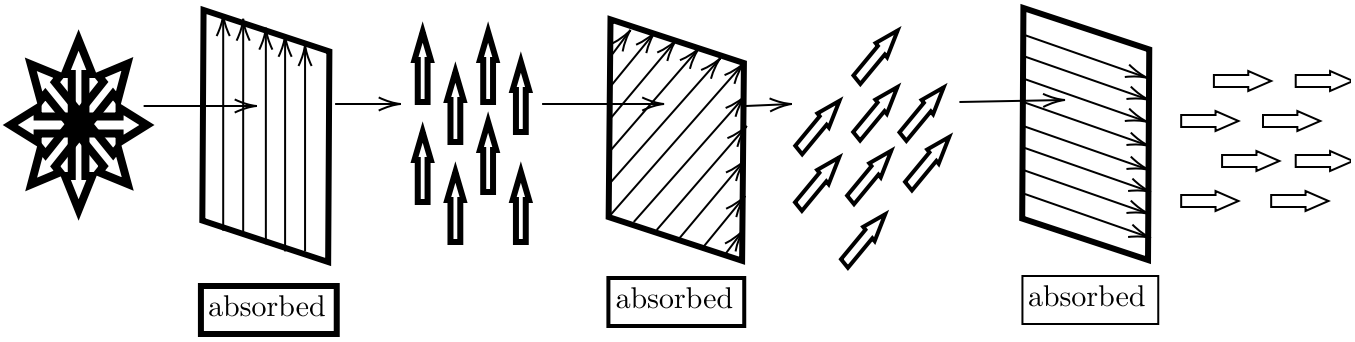


(2) But if the second device changes the orientation, then the particles once again behave nondeterministically with respect to the changed orientation. This is shown more cheaply using polarizing filters, except for not being able to identify the particles (of light) individually.



The individual photons do **not** "lose mojo" after their orientation "*collapses*" onto the basis state. It appears that way because of the physical fact that those photons giving the opposite outcome are absorbed by the filter.





In the second situation, the first filter produces light that is polarized up. The second filter absorbs $\cos^2\left(\frac{\pi}{4}\right) = \frac{1}{2}$ of that light and the other $\frac{1}{2}$ is passed through with diagonal polarization (analogous to the $|+\rangle$ basis state). The third filter absorbs $\frac{1}{2}$ again of that light. Positioning the middle filter at any angle θ between 0 and $\frac{\pi}{2}$ allows $\cos^2(\theta) \cdot \sin^2(\theta)$ of the light from the first filter to go through. This goes to zero as θ approaches either 0 or 90° and is maxed for $\theta = 45^\circ$. The Born Rule in action!

For most work with quantum circuits, we may suppose that a single measurement is taken at the end, and the output is read from the basis state $|y\rangle$ that is returned. Or we may run a circuit multiple times, thus **sampling** y from the output distribution. The **principle of deferred measurement**, which was seen in Chapter 6, makes this be "without loss of generality" in most computing situations---provided the measurement results are used only as controls for other gates. Quantum communication protocols, however, require a fuller formulation of measurement via linear algebra. This will come hand in hand with **mixed states**, which "are" classical probability distributions over unit vectors that are quantum **pure states**. Doing this is facilitated by the **Bloch Sphere** representation of qubits.

The Bloch Sphere

The previous (part of) lecture showed the limitations of the Cartesian picture for viewing even the simple computation $[a, b]^T = \mathbf{HTH}|0\rangle$. So we will study one that gives a different picture of physical reality.

The first point is that the complex numbers $a = x + iy$ and $b = u + iv$ involve 4 real numbers, but the requirement $|a|^2 + |b|^2 = 1$ imposes one constraint, thus essentially cutting the "real degrees of freedom" down to 3. A second factor cuts it down to 2. The following definition will be useful for quantum states of multiple qubits as well:

Definition: Two quantum states ϕ, ϕ' are **equivalent** if there is a unit complex number c such that $\phi' = c\phi$.

For example, $\frac{1}{\sqrt{2}}(-1, 1)$ is equivalent to $\frac{1}{\sqrt{2}}(1, -1)$, but neither is equivalent to $\frac{1}{\sqrt{2}}(1, 1)$, nor any of these to our basic states $(1, 0)$ and $(0, 1)$. In the line for the matrix $\mathbf{Y}$, $i\mathbf{e}_1$ is simply equivalent to just $\mathbf{e}_1$, $-i\mathbf{e}_0$ to $\mathbf{e}_0$, $-i\mu$ to μ, and $i\pi$. We could also regard $\mathbf{Y}$ as equivalent to

$$i\mathbf{Y} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

which makes clearer that it is a combination of $\mathbf{X}$ and $\mathbf{Z}$ (indeed, $i\mathbf{Y} = \mathbf{ZX} = -\mathbf{XZ}$). Finally, to finish the line for $\mathbf{Z}$, $\mathbf{Ze}_1 = -\mathbf{e}_1 \equiv \mathbf{e}_1$.

Regarding our saying *equivalence*, note that if $c = a + bi$, then

$$\frac{1}{c} = \frac{1}{a + bi} = \frac{a - bi}{(a + bi)(a - bi)} = \frac{a - bi}{a^2 + b^2} = \frac{a - bi}{1} = a - bi = \overline{c},$$

which is the **complex conjugate** of c and is likewise a unit complex number. Since $\phi = \overline{c}\phi'$ the relation is symmetric. That the product of two unit complex numbers is a unit complex number makes it transitive, and being reflexive is immediate with $c = 1$, so this is an equivalence relation.

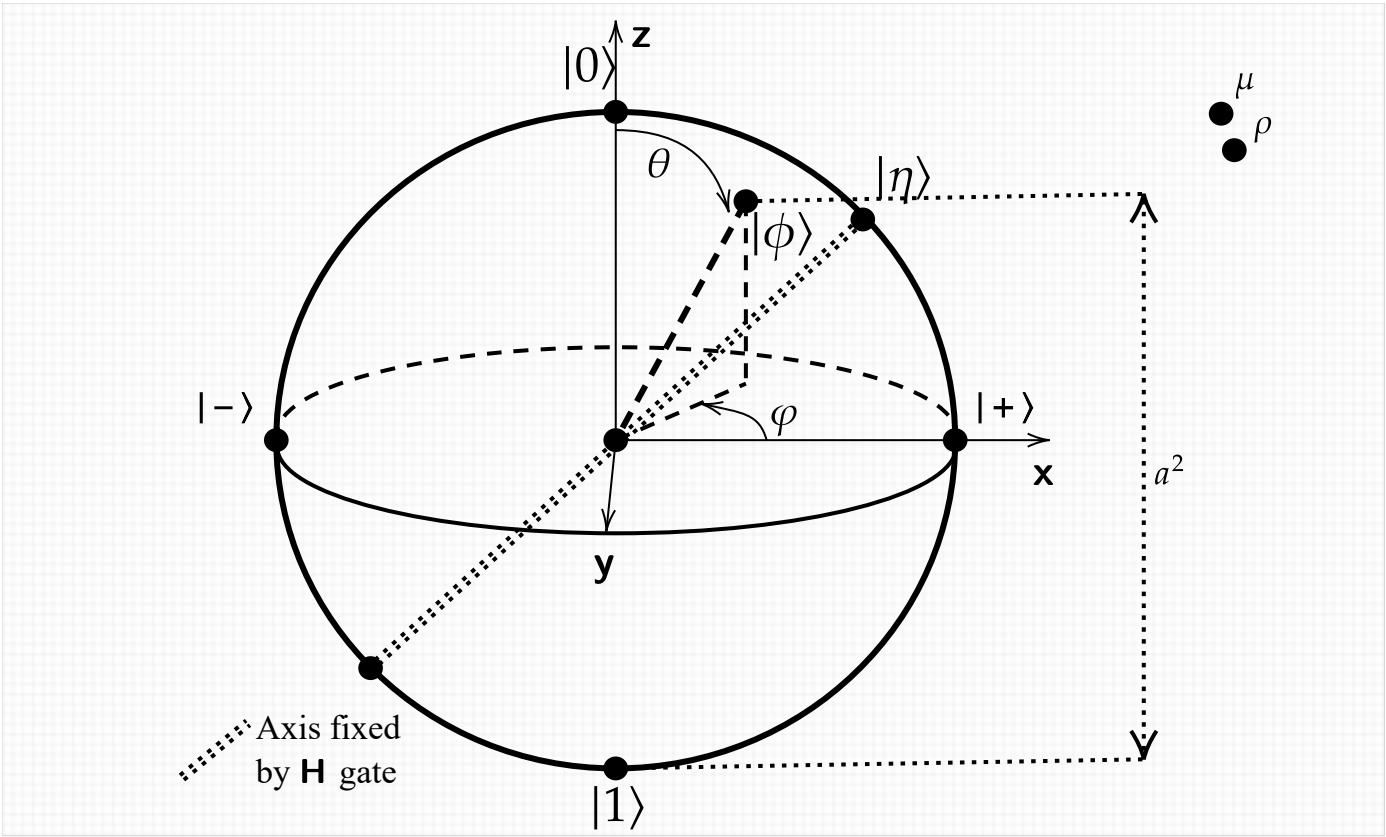
A unit complex number can be written in polar coordinates as $c = e^{i\gamma}$ for some angle γ, which represents a "global phase." Thus, dividing out by this equivalence relation emphasizes the **relative phase** φ of the two components. So let us write our original quantum state ϕ in polar coordinates as $\left(ae^{i\alpha}, be^{i\beta}\right)$ where now a, b are real numbers between 0 and 1. Choose $\gamma = -\alpha$, then $c\phi = \left(a, be^{i\varphi}\right)$ with $\varphi = \beta - \alpha$. Since $a^2 + b^2 = 1$, the value of b is forced once we specify a. So a and φ are enough to specify the state. These are the 2 true degrees of freedom.

We can uniquely map points (a, φ) to the sphere by treating φ as a longitude and a^2 (rather than a) as a latitude where the north pole is 1, the equator is 0.5, and the south pole is 0. Then the latitude gives

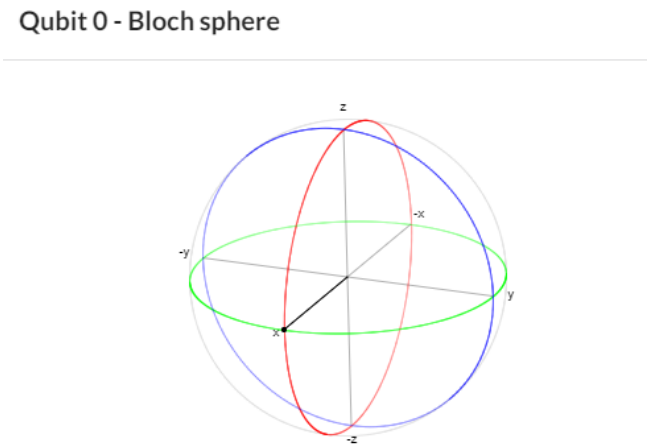
the probability of getting the outcome 0. All states that give equal probability of 0 and 1 fan out along the equator. The north pole is $|0\rangle$ and the south pole is $|1\rangle$. And again:

- $\frac{1}{\sqrt{2}}(1, 1) = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)$ is called $|+\rangle$, the "plus" state.
- $\frac{1}{\sqrt{2}}(1, -1) = \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right)$ is called $|-\rangle$, the "minus" state.

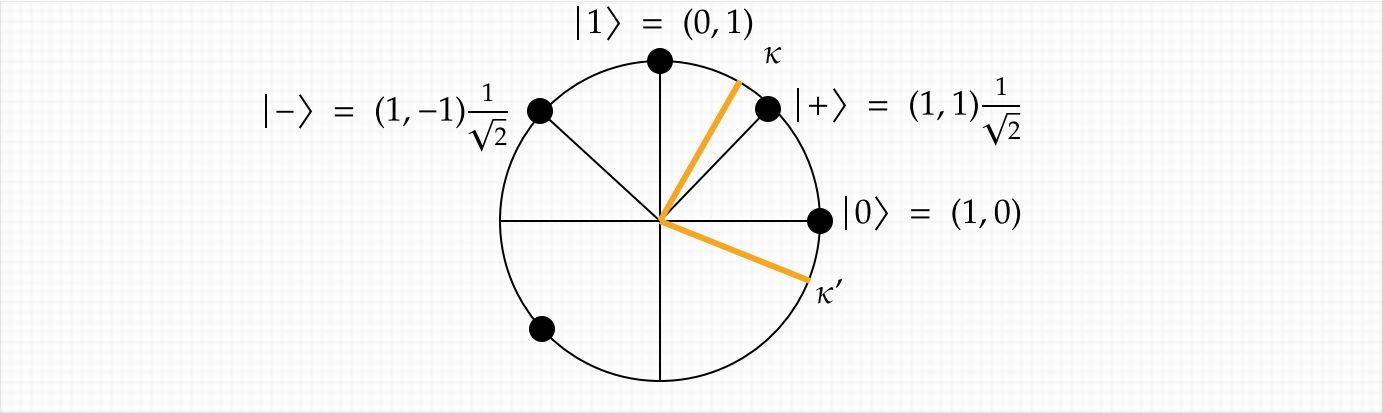
Here they all are, graphed on the **Bloch Sphere**:



Among web applets displaying Bloch spheres for qubits is https://quantum-circuit.com/home (free registration required). Here is its graph for the $|+\rangle$ state. It is more usual to show the $\mathbf{x}$ axis out toward the reader and $\mathbf{y}$ at right, but that is less convenient IMHO for picturing $|+\rangle$ and $|-\rangle$.



Some algorithms, however, are IMHO easier to picture using the original planar diagram in Cartesian coordinates:



For one thing, this makes it easier to tell that $|0\rangle$ and $|1\rangle$ are orthogonal vectors, that $|+\rangle$ and $|-\rangle$ are likewise orthogonal vectors, and that the orthonormal basis $\{|+\rangle, |-\rangle\}$ is obtained by a linear transformation (indeed, a simple rotation) of the standard basis $\{|0\rangle, |1\rangle\}$. We will even use this to illustrate the **CHSH Game**.

A downside, however, is that this diagram gives extra points for equivalent space, whereas the Bloch sphere is completely non-redundant. The Bloch sphere is also "more real" than the way we usually graph complex numbers via Cartesian coordinates. In fact, *every unitary* 2×2 *matrix* U *induces a rotation of the Bloch sphere and hence fixes an axis, so the axes of the sphere are in 1-to-1 correspondence with lossless quantum operations on a single qubit.* Whereas, the planar diagram gives a cut-down picture of how $\mathbf{H}$ acts as a rotation without fully showing you its axis.

The axis of the $\mathbf{H}$ gate goes through the origin and the point corresponding to the pure state $|\eta\rangle = \left[cos\frac{\pi}{8}, \sin\frac{\pi}{8}\right]$. With this vector, the latitude is $\cos^2\left(\frac{\pi}{8}\right) = 0.85355339...$ That's the number we got from the $\mathbf{HTH}$ computation. Note: the latitude looks like it should be "3/4" but it's not. The equator is 0.5 and the diagonal point is $\frac{1}{\sqrt{2}}$ of the way up from equator to the pole, so the latitude is $0.5 + 0.5\frac{1}{\sqrt{2}} = 0.85355339...$ as required.

Mixed States and Quantum States as Operators

A **pure state** of n qubits is one denoted by a unit vector in $\mathbb{C}^{2^n}$. A **mixed state** is any linear combination of pure states by non-negative weights that sum to 1. That is, a mixed state is a **classical** probability distribution over pure states. Whether "mixed state" includes pure states depends on context; one can say "properly mixed" to exclude pure states.

For one qubit, every properly mixed state maps to a point interior to the Bloch Sphere. This also holds for generalizations of the Bloch Sphere to higher dimensions for more qubits. So let us have pure states $|\phi_1\rangle, \ldots, |\phi_m\rangle$ and probabilities $p_1, \ldots, p_m$ summing to 1. Then

$$p_1|\phi_1\rangle + \cdots + p_m|\phi_m\rangle$$

is the "standard" representation of the mixed state. We will see momentarily that, like writing $|\phi_k\rangle$ to begin with, it may presume more than we can directly sense. A philosophical question that comes first is whether a mixed state is a "thing", or just our lack of full knowledge about the state. To appreciate this, we need to treat both pure and mixed states as operators and formalize more about how measurements are represented in any basis.

Definition: For any mixed state represented as $p_1|\phi_1\rangle + p_2|\phi_2\rangle + \cdots + p_m|\phi_m\rangle$, where the p_i are nonnegative and sum to 1, the corresponding **density matrix** is

$$\rho = p_1|\phi_1\rangle\langle\phi_1| + p_2|\phi_2\rangle\langle\phi_2| + \cdots + p_m|\phi_m\rangle\langle\phi_m|.$$

Per the above philosophy, ρ is all we can know about the mixed state (aside from any prior knowledge from having prepared it). The letter ρ tends to be used, without a ket or bra around it. Some more facts:

1. Since it is a weighted sum of outerproducts, a density matrix is always Hermitian: $\rho^* = \rho$.
2. The matrix designates a pure state if and only if $\rho^2 = \rho$; note that this is automatic as shown above when $m = 1$.
3. The results of measuring a mixed state can be computed by applying ρ as an operator to update the state. By linearity, this is the same as working with each individual term and taking the linear combination.

For example, the mixed state obtained by averaging the two basis states is

$$0.5|0\rangle\langle 0| + 0.5|1\rangle\langle 1| = 0.5\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 0.5\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

This is not the same as $|+\rangle\langle +| = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$. Note that the square of the former martrix is $\begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}$, which is not the same and no longer has **trace** equal to 1. The **trace** $Tr(M)$ of a square matrix M is the sum of the entries on the main diagonal. (We will later see a related notion for non-square matrices.) Whereas, the square of the latter matrix is itself.

[Lecture on 11/6/25 ended here. Tuesday 11/11 will pick up with more examples of mixed states.]