CSE439 Fall 2025 Week 12: Measurements, Spectral Theorem, and the CHSH Game

The mantra becomes: Everything is an Operator. Not just transformations, but quantum states and measurements too. We cannot do all this with our seminal notion of unitary matrices. The concept of Hermitian matrices, however, fills all the billings. Via the Spectral Theorem, we will see that unitary and Hermitian matrices have a deep connection via eigenvalues.

Examples of Mixed States

We can consider any probability mixture of the $|0\rangle$ and $|1\rangle$ basis states. The density matrix of the mixed state $p|0\rangle + (1-p)|1\rangle$ is

$$\boldsymbol{\rho}_p \;=\; p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| \;=\; p\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + (1-p)\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix}.$$

Note that $\boldsymbol{\rho}_p^2 = \begin{bmatrix} p^2 & 0 \\ 0 & (1-p)^2 \end{bmatrix} \neq \rho_p$ unless $p = 1$ or $p = 0$, so this is generally not a pure state.

How about $p|+\rangle\langle +| \;+\; (1-p)|-\rangle\langle -|$? We get

$$\frac{1}{2}\left(p\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + (1-p)\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\right) = \frac{1}{2}\left(\begin{bmatrix} p & p \\ p & p \end{bmatrix} + \begin{bmatrix} 1-p & p-1 \\ p-1 & 1-p \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} 1 & 1-2p \\ 1-2p & 1 \end{bmatrix}.$$

In general, this is different. But for the equal mixture $p = \frac{1}{2}$, both density matrices are the same: $\boldsymbol{\rho}_{1/2} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$. In terms of the Bloch sphere, both mixtures map to the exact center of the sphere, which is halfway down the axis between $|0\rangle$ and $|1\rangle$ at the poles, and also halfway along the equatorial axis between $|+\rangle$ and $|-\rangle$. In physical terms, that means they are *the same state*. That might come as a surprise, because:

> One is defined as a spread between the outcomes $|0\rangle$ and $|1\rangle$, the other between the outcomes $|+\rangle$ and $|-\rangle$. Isn't that like saying one is apple vs. pear, the other orange vs. grapefruit?

The ultimate point is that to probe the state, we have to choose a basis to measure against in advance. If we choose the standard basis, then to measure the probability for the outcome $|0\rangle$, even if we use the $|+\rangle$ and $|-\rangle$ mixture, we still get

$$P_{|0\rangle}(\boldsymbol{\rho}_{1/2}) \;=\; \langle 0|\big(0.5|+\rangle\langle +| \;+\; 0.5|-\rangle\langle -|\big)|0\rangle \;=\; 0.5\langle 0|+\rangle\langle +|0\rangle \;+\; 0.5\langle 0|-\rangle\langle -|0\rangle$$
$$= \; 0.5\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}} \;+\; 0.5\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}} \;=\; 0.5.$$

Note that this associated the terms so that the fact that the $|0\rangle$ and $|+\rangle$ vectors are 45° aligned to each other in Cartesian coordinates, likewise $|0\rangle$ and $|-\rangle$, came out as an idea. But we can get the point much more succinctly upon measuring any outcome $|\kappa\rangle$ for $\boldsymbol{\rho}_{1/2}$:

$$\langle\kappa|\boldsymbol{\rho}_{1/2}|\kappa\rangle \;=\; \left\langle\kappa\middle|\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\middle|\kappa\right\rangle \;=\; \langle\kappa|0.5\mathbf{I}|\kappa\rangle \;=\; 0.5\langle\kappa|\mathbf{I}|\kappa\rangle \;=\; 0.5\langle\kappa|\kappa\rangle \;=\; 0.5.$$

That's it. However we try to probe the **completely mixed state** $\boldsymbol{\rho}_{1/2}$, it just behaves like a perfect unbiased classical coin. Regardless of what we mixed to make it, there is nothing else that it is now.

General Measurements and Operators [My 11/11/25 lecture cut down the parts in brown]

The **triple product** of a row-vector x, a matrix A, and a column vector y is just xAy. We will care about the case where x is the "bra" dual of y and A is an outer-product matrix (or some other Hermitian matrix). Let's write $y = |\kappa\rangle = [a, b]^T$, where a and b are complex numbers such that $|a|^2 + |b|^2 = 1$. Now consider the fact that the inner product of $[1, 0]$ with $|\kappa\rangle$, i.e., of $|0\rangle$ but written it as the bra $\langle 0|$, is just a. Meanwhile the inner product $\langle\kappa|\cdot|0\rangle$ gives a^*. Furthermore,

$$a^*a \;=\; \langle\kappa|\cdot|0\rangle\langle 0|\cdot|\kappa\rangle \;=\; \langle\kappa|0\rangle\cdot\langle 0|\kappa\rangle \;=\; |\langle\kappa|0\rangle|^2 \;=\; |a|^2.$$

What this says is that we **projected** the vector denoted by κ **onto** the basis vector $|0\rangle$, and then took the magnitude of that projection. Thus $|0\rangle\langle 0|$ represents the operation of **projecting onto the** $|0\rangle$ **vector**. Moreover, look how it transforms the $|\kappa\rangle$ vector:

$$\big(|0\rangle\langle 0|\big)\cdot|\kappa\rangle \;=\; |0\rangle\cdot\langle 0|\kappa\rangle \;=\; |0\rangle(1\cdot a + 0\cdot b) \;=\; a|0\rangle.$$

If we let $p_0 = |a|^2$ stand for the probability of $|0\rangle$ and divide through by $\sqrt{p_0}$ then we get just $|0\rangle$. Oh wait, what we actually get is

$$\frac{1}{\sqrt{p_0}}\big(|0\rangle\langle 0|\big)\cdot|\kappa\rangle \;=\; \frac{1}{\sqrt{p_0}}a|0\rangle \;=\; \frac{a}{|a|}|0\rangle.$$

This might not be exactly $|0\rangle$, but it is **equivalent** to it since $\dfrac{a}{|a|}$ is always a unit complex scalar. That's good enough. Thus $\dfrac{1}{\sqrt{p_0}}\big(|0\rangle\langle 0|\big)$ updates the state when outcome $|0\rangle$ happens. Similarly, $\dfrac{1}{\sqrt{p_1}}\big(|1\rangle\langle 1|\big)$ faithfully updates the state when outcome $|1\rangle$ happens. Again, the point is how this works for any basis, not just the standard basis. Let's see the general definitions first, then do things within the $|+\rangle, |-\rangle$ basis, then use $|+\rangle, |-\rangle$ to measure κ as originally defined via $a|0\rangle + b|1\rangle$.

Definition: The **projection operator** associated to a pure state $|\phi\rangle$ is $\mathbf{P}_\phi = |\phi\rangle\langle\phi|$.

Note that $\mathbf{P}_\phi^* = \left(|\phi\rangle \cdot \langle\phi|\right)^* = \langle\phi|^* \cdot |\phi\rangle^* = |\phi\rangle \cdot \langle\phi| = \mathbf{P}_\phi$, so every projection operator is Hermitian. More generally, we define:

Definition: A matrix B is **positive semidefinite** (PSD) if there is a matrix A such that $B = AA^*$.

Definition: A matrix P computes a **projection** if it is PSD and $P^2 = P$.

By $\mathbf{P}_\phi^* = \mathbf{P}_\phi$ we also have

$$\mathbf{P}_\phi \mathbf{P}_\phi^* = \mathbf{P}_\phi^2 = |\phi\rangle\langle\phi| \cdot |\phi\rangle\langle\phi| = |\phi\rangle \cdot \langle\phi|\phi\rangle \cdot \langle\phi| = |\phi\rangle \cdot 1 \cdot \langle\phi| = \mathbf{P}_\phi,$$

since $|\phi\rangle$ is a unit vector. So $\mathbf{P}_\phi$ is indeed a projection and is PSD.

Definition: A **projective measurement** is given by a set $\{\mathbf{P}_1, \ldots, \mathbf{P}_m\}$ of projections such that

$$\sum_{i=1}^{m} \mathbf{P}_i = \mathbf{I}.$$

From above, $\left\{|0\rangle\langle 0|, |1\rangle\langle 1|\right\}$ is a projective measurement. How about the $\mathbf{X}$ basis $\left\{|+\rangle\langle +|, |-\rangle\langle -|\right\}$? Using the numerics of the standard basis, we get:

$$|+\rangle\langle +| = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^T \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right] = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$|-\rangle\langle -| = \left[\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right]^T \left[\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right] = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

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

So $\left\{|+\rangle\langle +|, |-\rangle\langle -|\right\}$ is a projective measurement. Note that if we used the $|+\rangle, |-\rangle$ coordinates to begin with, then the numerics would be $|+\rangle\langle +| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and would come out literally identical, likewise if we apply the measurement to $|\kappa'\rangle = a|+\rangle + b|-\rangle$. (Note: the third from last line on page 145 would be less confusing if it defined $|\kappa'\rangle$ this way rather than say $|\kappa\rangle$ again.) Using the standard-basis numerics:

$$|\kappa'\rangle = \left[\frac{a}{\sqrt{2}}, \frac{a}{\sqrt{2}}\right]^T + \left[\frac{b}{\sqrt{2}}, \frac{-b}{\sqrt{2}}\right]^T = \frac{1}{\sqrt{2}}[a+b, a-b]^T.$$

The triple product with $|+\rangle\langle+|$ is:

$$\langle\kappa'|\cdot|+\rangle\langle+|\cdot|\kappa'\rangle = \frac{1}{4}\left[a^*+b^*, a^*-b^*\right]\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}[a+b, a-b]^T = \frac{1}{4}\left[2a^*, 2a^*\right]\begin{bmatrix}a+b\\ a-b\end{bmatrix}$$
$$= \frac{1}{4}\left(2a^*a + 2a^*b + 2a^*a - 2a^*b\right) = \frac{1}{4}\left(4a^*a\right) = a^*a = |a|^2.$$

Similarly, we get $\langle\kappa'|\cdot|-\rangle\langle-|\cdot|\kappa'\rangle = |b|^2$. That is a lot of rigamarole to replicate the answer we got for measuring the original $|\kappa\rangle$ in the standard basis. The larger point is that the $|\kappa'\rangle$ vector with regard to the **X** basis has the same relation to it as $|\kappa\rangle$ did to the standard basis.

However, when we expressly write $|\kappa\rangle = a|0\rangle + b|1\rangle$ rather than $|\kappa\rangle = [a, b]^T$, then we are defining it in a way that is independent of a particular coordinate notation, and so it really is a different physical vector from $|\kappa'\rangle = a|+\rangle + b|-\rangle$. To underscore the point (this is an example that should be on page 146), let us measure $|\kappa\rangle$ not $|\kappa'\rangle$ in the **X** basis.

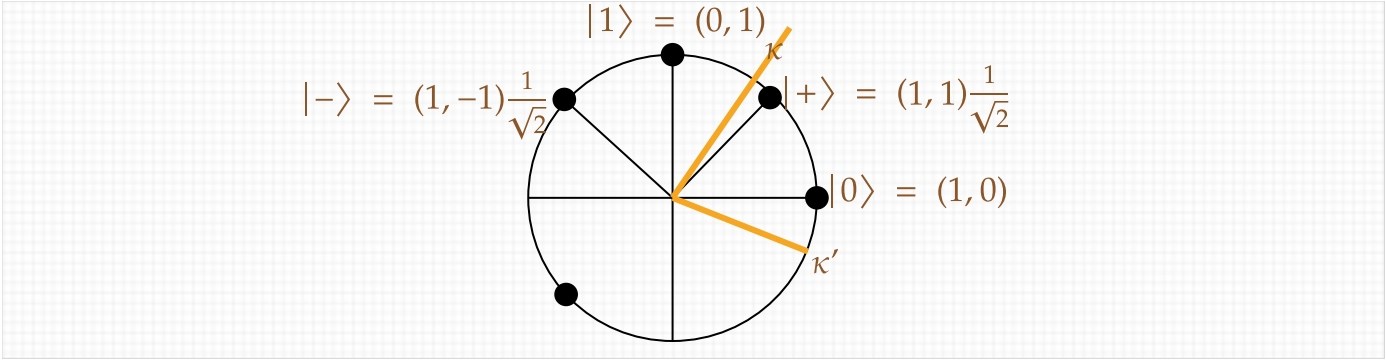
$$\langle\kappa|\cdot|+\rangle\langle+|\cdot|\kappa\rangle = \left[a^*, b^*\right]\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}[a, b]^T = \frac{1}{2}\left[a^*+b^*, a^*+b^*\right]\begin{bmatrix}a\\ b\end{bmatrix}$$
$$= \frac{1}{2}\left(a^*a + a^*b + b^*a + b^*b\right) = \frac{1}{2}\left(|a|^2 + |b|^2 + a^*b + b^*a\right) = \frac{1}{2} + \frac{c+c^*}{2}$$

where $c = a^*b$. What happened? The first thing to note is that the sum of a unit complex number c and its conjugate is always a real number because the imaginary parts cancel. Although in general the sum could be as big as 2 (or as low as -2), because c arises as a^*b where $|a|^2 + |b|^2 = 1$, the maximum magnitude of $c + c^*$ is 1. Hence the probability of getting the outcome $|+\rangle$ stays within the range $[0, 1]$ as required for a probability.

In fact, if $\kappa = |+\rangle$ then $c = \frac{1}{2}$ and $c + c^* = 1$, finally giving that the probability of getting the outcome $|+\rangle$ is 1. And the probability of getting the outcome $|-\rangle$ is:

$$\langle\kappa|\cdot|-\rangle\langle-|\cdot|\kappa\rangle = \left[a^*, b^*\right]\frac{1}{2}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}[a, b]^T = \frac{1}{2}\left[a^*-b^*, -a^*+b^*\right]\begin{bmatrix}a\\ b\end{bmatrix}$$
$$= \frac{1}{2}\left(a^*a - b^*a - a^*b + b^*b\right) = \frac{1}{2}\left(|a|^2 + |b|^2 - a^*b - b^*a\right) = \frac{1}{2} - \frac{c+c^*}{2}$$

with $c = a^*b$ as before. This ensures that the probabilities sum to 1, regardless of what c is. It is a nice self-study exercise to repeat this with the example $|\kappa\rangle = \left[\frac{1}{2}, \frac{\sqrt{3}}{2}\right]$.



There is an essential symmetry of measurement as well. If we instead did $\langle -|\cdot|\kappa\rangle\langle\kappa|\cdot|-\rangle$ then we would get the same answer. Indeed, for a general other pure state $|\phi\rangle$, the **double action**

$$P_{|\kappa\rangle}(|\phi\rangle) = \langle\phi|\cdot|\kappa\rangle\langle\kappa|\cdot|\phi\rangle$$

is a product of the form cc^* where $c = \langle\phi|\kappa\rangle$. And $(cc^*)^* = (c^*)^*(c)^* = cc^*$ back again, so the *product* of a complex number and its conjugate is always a real number too. Some interpretations:

- The only knowledge we can *gain* about a quantum state $|\kappa\rangle$ (relative to any prior knowledge about how it was prepared) is by *measuring* it.
- All measurements of κ go through the outer product $|\kappa\rangle\langle\kappa|$.
- Hence $|\kappa\rangle\langle\kappa|$, not $|\kappa\rangle$, is the "unit of epistemology" (the origin of "episte-" is the idea of sending a message, i.e., an *epistle*). This is a Hermitian operator and a PSD matrix with real entries and a projection. All complex numbers have vamoosed.

This carries through when $|\kappa\rangle$ is a state of multiple qubits, or of multiple **qutrits**, **quarts**, **qudits** (meaning d-ary, as with card ranks where $d = 13$), and so on, even going into infinite-dimensional Hilbert spaces. The "real proof" of the principle, IMHO, comes from the extension to mixed states.

The Spectral Theorem

Theorem (split between theorems 14.1 and 18.1 in the text): If A is an $n \times n$ Hermitian matrix, then there are n real numbers $\lambda_1, \ldots, \lambda_n$ (not necessarily all distinct) and associated vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ forming an orthonormal basis, such that

$$A = \lambda_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \lambda_2|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \lambda_n|\mathbf{u}_n\rangle\langle\mathbf{u}_n|.$$

Furthermore, the matrix e^{iA}, which is then well-defined by

$$e^{iA} = e^{i\lambda_1}|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + e^{i\lambda_2}|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + e^{i\lambda_n}|\mathbf{u}_n\rangle\langle\mathbf{u}_n|,$$

is unitary. And for the converse: *every* unitary matrix arises in this manner.

Proof: The first part is by induction. By the fundamental theorem of algebra, the characteristic polynomial $det(A - xI)$ has n solutions over $\mathbb{C}$, counting multiplicities. If there is only one distinct solution λ, then A must equal λI. By the Hermitian property $A^* = A$, λ must be real, and we can get $A = \lambda I = \lambda|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \lambda|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \lambda|\mathbf{u}_n\rangle\langle\mathbf{u}_n|$ from any orthonormal basis of the space. This is the base case. Note also that for $n = 1$, the basis is unique.

So suppose λ_1 is one of at least two distinct solutions. Then the subspace W of vectors $\mathbf{v}$ such that $A\mathbf{v} = \lambda_1\mathbf{v}$ is not the whole space---it has dimension m less than n. So let $\mathbf{x}$ be in W and $\mathbf{y}$ in the orthogonal complement $W^\perp$ of W. By the Hermitian property,

$$\langle x, Ay\rangle = \langle Ax, y\rangle = \langle \lambda_1^* x, y\rangle = \lambda_1^*\langle x, y\rangle = 0.$$

Since x is an arbitrary vector in W, this means that Ay always stays in the orthogonal complement $W^\perp$, as well as Ax always staying within W. Hence we can argue inductively about A acting on W and on $W^\perp$ individually. This induction also concludes, as ultimately validated on hitting the base case, that λ_1 is real, so $\lambda_1^* = \lambda_1$, and this carries through to all other (distinct) solutions. This process also builds othonormal vectors $\mathbf{u}_i$ such that

$$A = \lambda_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \lambda_2|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \lambda_n|\mathbf{u}_n\rangle\langle\mathbf{u}_n|.$$

Note that these are automatically eigenvectors, because

$$\begin{aligned} A\mathbf{u}_i &= \lambda_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1|\mathbf{u}_i + \lambda_2|\mathbf{u}_2\rangle\langle\mathbf{u}_2|\mathbf{u}_i + \cdots + \lambda_i|\mathbf{u}_i\rangle\langle\mathbf{u}_i|\mathbf{u}_i + \lambda_n|\mathbf{u}_n\rangle\langle\mathbf{u}_n|\mathbf{u}_i \\ &= \lambda_1 |\mathbf{u}_1\rangle \cdot 0 + \lambda_2 |\mathbf{u}_2\rangle \cdot 0 + \cdots + \lambda_i |\mathbf{u}_i\rangle \cdot 1 + \cdots + \lambda_n |\mathbf{u}_n\rangle \cdot 0 \\ &= \lambda_i\mathbf{u}_i \end{aligned}$$

(Well, this is because the notation $|\mathbf{u}_i\rangle$ and just $\mathbf{u}_i$ is interchangeable.) Moreover, if λ_i has multiplicity 1, i.e. is a unique eigenvalue in its eigenspace, then the associated unit eigenvector $\mathbf{u}_i$ is unique. Now to show that e^{iA} is unitary, we note that its adjoint is

$$\overline{e^{iA}}^T = e^{-iA^T} = e^{-i\lambda_1}|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + e^{-i\lambda_2}|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + e^{-i\lambda_n}|\mathbf{u}_n\rangle\langle\mathbf{u}_n|.$$

This is because, as we've seen, every self-outerproduct $|\mathbf{u}\rangle\langle\mathbf{u}|$ is Hermitian so those parts don't change under conjugate transpose. Finally, when we multiply e^{iA} by its adjoint, all of the cross-terms cancel by the orthogonality of the $\mathbf{u}_i$ vectors, leaving only the products of like terms:

$$e^{i\lambda_1}|\mathbf{u}_1\rangle\langle\mathbf{u}_1|e^{-i\lambda_1}|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \cdots + e^{i\lambda_n}|\mathbf{u}_n\rangle\langle\mathbf{u}_n|e^{-i\lambda_n}|\mathbf{u}_n\rangle\langle\mathbf{u}_n|$$

$$= e^{i\lambda_1}e^{-i\lambda_1}|\mathbf{u}_1\rangle\langle\mathbf{u}_1||\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \cdots + e^{i\lambda_n}e^{-i\lambda_n}|\mathbf{u}_n\rangle\langle\mathbf{u}_n||\mathbf{u}_n\rangle\langle\mathbf{u}_n|$$

$$= |\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \cdots + |\mathbf{u}_n\rangle\langle\mathbf{u}_n| = I,$$

because $e^{i\lambda_1}e^{-i\lambda_1} = e^{i(\lambda_1-\lambda_1)} = e^0 = 1$ (etc.) and the $\mathbf{u}_i$ are unit vectors. So e^{iA} is unitary.

For the converse direction, let U be any unitary matrix, and put

$$V = \frac{1}{2}(U + U^*) \text{ and } W = \frac{1}{2i}(U - U^*),$$

so that $U = V + iW$. These are intuitively trying to be the real and imaginary parts of the matrix U. Partial success is attested by the fact that they are Hermitian: $V^* = V$ and $W^* = W$. Moreover, $VW = WV$ because UU^* and U^*U both equal I.

Now a useful fact: Hermitian matrices A, B that commute can have the same orthonormal eigenbasis. For intuition, suppose λ_i has multiplicity 1 for A with unique unit eigenvector u_i. Take $v_i = Bu_i$. Then $Av_i = ABu_i = BAu_i = B\lambda_i u_i = \lambda_i Bu_i = \lambda_i v_i$. Thus v_i is also an eigenvector of A. It need not be a unit eigenvector like u_i, but it must be a multiple of u_i because the eigenspace is one-dimensional. So $Bu_i = v_i = \mu_i u_i$ for some constant μ_i. This constant can be different from λ_i, but it is an eigenvalue of B for the same eigenvector u_i. The general case of higher multiplicity is messier---and it is not the case that *every* orthonormal eigenbasis for A becomes one for B, only that *some* orthonormal eigenbasis of A carries over to B---but the basic reason it works is similar. Therefore, we can write:

$$\begin{aligned} V &= \lambda_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \lambda_2|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \lambda_n|\mathbf{u}_n\rangle\langle\mathbf{u}_n| \text{ and} \\ W &= \mu_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \mu_2|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \mu_n|\mathbf{u}_n\rangle\langle\mathbf{u}_n| \end{aligned}$$

with different eigenvalues λ_i, μ_i but the same vectors u_i. So

$$U = V + iW = (\lambda_1 + i\mu_1)|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + (\lambda_2 + i\mu_2)|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + (\lambda_n + i\mu_n)|\mathbf{u}_n\rangle\langle\mathbf{u}_n|.$$

Thus each $(\lambda_j + i\mu_j)$ is an eigenvalue of U. Since U is unitary, its eigenvalues have norm 1. Thus λ_j and μ_j are real numbers whose squares sum to 1, and they are therefore the cosine and sine of some angle θ_j. So

$$\lambda_j + i\mu_j = \cos\theta_j + i\sin\theta_j = e^{i\theta j}.$$

This means that we get a Hermitian matrix A such that $U = e^{iA}$ by taking

$$A = \theta_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \theta_2|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \theta_n|\mathbf{u}_n\rangle\langle\mathbf{u}_n|. \qquad \boxtimes$$

Numerical Matrix Operations

One major application of the spectral representation of a matrix A (when A is Hermitian so it is available) is in representing and executing numerical functions $f(x)$ as matrix functions $f(A)$. We have seen this already with $f(x) = e^{ix}$ as defining "phased exponentiation" e^{iA}. This can be defined in general given $A = \lambda_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \lambda_2|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \lambda_n|\mathbf{u}_n\rangle\langle\mathbf{u}_n|$:

$$f(A) = f(\lambda_1)|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + f(\lambda_2)|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + f(\lambda_n)|\mathbf{u}_n\rangle\langle\mathbf{u}_n|.$$

When f is a function involving addition and subtraction and multiplication only (i.e., is a *polynomial function*) then this is immediately evident: only multiplication needs a second thought, and it works because terms for different orthogonal eigenvectors $\mathbf{u}_i$,$\mathbf{u}_j$ will cancel when multiplied. Provided A and B are decomposed in the same eigenbasis, this works for two-variable functions $f(x, y)$ as well: if

$$A = \lambda_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \lambda_2|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \lambda_n|\mathbf{u}_n\rangle\langle\mathbf{u}_n| \text{ and}$$
$$B = \mu_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \mu_2|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \mu_n|\mathbf{u}_n\rangle\langle\mathbf{u}_n|$$

then

$$f(A, B) = f(\lambda_1, \mu_1)|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + f(\lambda_2, \mu_2)|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + f(\lambda_n\mu_n)|\mathbf{u}_n\rangle\langle\mathbf{u}_n|.$$

But the fun is that this works for just about any function f. *A fortiori*, this is because just about any function is approximable by polynomials. For example,

$$A^{-1} = \frac{1}{\lambda_1}|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \frac{1}{\lambda_2}|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \frac{1}{\lambda_n}|\mathbf{u}_n\rangle\langle\mathbf{u}_n|$$

(Wait a second---we saw that many Hermitian matrices, including ones from outer-products $|\phi\rangle\langle\phi|$, are **not** invertible. So how can we do this?? Well, those matrices have 0 as an eigenvalue occurring at least once. So the above definition would try to do $1/0$, which blows up. So no contradiction here.)

This idea, plus using a polynomial approximation to the numerical function $1/x$ that works on a needed interval bounded away from $x = 0$, is the jumping point for the **HHL Algorithm** for (approximately) solving matrix equations by (*approximate*) inversion, as covered in Chapter 18. Another example is:

$$\sqrt{A} = \sqrt{\lambda_1}\cdot|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \sqrt{\lambda_2}\cdot|\mathbf{u}_2\rangle\langle\mathbf{u}_2| + \cdots + \sqrt{\lambda_n}\cdot|\mathbf{u}_n\rangle\langle\mathbf{u}_n|.$$

For unitary matrices A that happen to also be Hermitian, such as the Pauli matrices and **CNOT** and **CZ**, this gives a way to compute square roots for them. For example we can represent **CNOT** as:

$$1\cdot\left|\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\right\rangle\left\langle\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\right| + 1\cdot\left|\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\right\rangle\left\langle\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\right| + 1\cdot\left|\begin{bmatrix}0\\0\\1\\1\end{bmatrix}\right\rangle\left\langle\begin{bmatrix}0\\0\\1\\1\end{bmatrix}\right|\frac{1}{2} + (-1)\cdot\left|\begin{bmatrix}0\\0\\1\\-1\end{bmatrix}\right\rangle\left\langle\begin{bmatrix}0\\0\\1\\-1\end{bmatrix}\right|\frac{1}{2}.$$

To get a 4×4 matrix B such that $B^2 = \mathbf{CNOT}$ we just take square roots of all the eigenvalues. We have a wide choice: $+1$ or -1 for the first there and i or $-i$ for the -1. Using the positive signs gives

$$B = 1 \cdot \left| \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\rangle \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right| + 1 \cdot \left| \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\rangle \left\langle \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right| + 1 \cdot \left| \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\rangle \left\langle \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right| \frac{1}{2} + i \cdot \left| \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right\rangle \left\langle \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right| \frac{1}{2}$$

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

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

This is the matrix of the controlled gate $\mathbf{CV}$ where $\mathbf{V} = \frac{1}{2} \begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}$ is also written $\mathbf{X}^{\frac{1}{2}}$ and called the **square root of NOT**. (The Wybiral circuit simulator calls it "SRNOT".) Notice also that $\mathbf{V}$ is not Hermitian like $\mathbf{X}$ is---but that's OK since i is not an eigenvalue of this basis (nor a real number, either).

[The 11/11/25 lecture basically got up to here; the converse of the Spectral Theorem was on Thu. 11/13. In that lecture, I did one more example, which follows next.]

Another Example: Let $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$. This is Hermitian but not unitary---in fact, it's not invertible. It's not normalized in the sense of having $Tr(A) = 1$ either, though if we divide it by 2 then it does become $|-\rangle\langle -|$. Let's leave it un-normalized on purpose to see what happens. Since it's not invertible---indeed has rank one less than **full rank**---it follows that 0 occurs once as an eigenvalue. The characteristic polynomial $\det(xI - A)$ [I usually use $\det(A - xI)$ to make fewer typos] is $(x-1)^2 - 1 = x^2 - 2x$ so the other eigenvalue is $+2$. It's easy to see that $[1, 1]^T$ is an eigenvector of 0 and $A \cdot [{}^1_{-1}] = [{}^2_{-2}]$ so that $[1, -1]^T$ is an eigenvector of 2. These eigenvectors are also orthogonal. We do need to normalize the eigenvectors, so we get

$$\mathbf{u_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = |+\rangle \text{ and } \mathbf{u_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = |-\rangle$$

after all. Then

$$A = 0 \cdot |\mathbf{u_1}\rangle\langle \mathbf{u_1}| + 2 \cdot |\mathbf{u_2}\rangle\langle \mathbf{u_2}| = 0 \cdot \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 2 \cdot \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

It is tempting to simplify the constants, but we want to keep the actual eigenvalues separate. To get a square root of A, the second term needs to have $\sqrt{2} \cdot \frac{1}{2}[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}]$ not $\sqrt{1} \cdot [\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}]$. That works since $A^2 = 2A$, so taking $B = \frac{1}{\sqrt{2}}A$ makes $B^2 = \frac{1}{2}A^2 = \frac{1}{2} \cdot 2A = A$. To get the unitary matrix e^{iA}, we need to keep the 0 eigenvalue separate too:

$$e^{iA} \;=\; e^{i\cdot 0} \cdot \tfrac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \;+\; e^{i\cdot 2} \cdot \tfrac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \;=\; \tfrac{1}{2}\begin{bmatrix} 1+e^{2i} & 1-e^{2i} \\ 1-e^{2i} & 1+e^{2i} \end{bmatrix}.$$

What the heck is e^{2i} anyway?? Well, it's $e^{i\theta}$ where $\theta = 2$. Not 2 degrees, but 2 *radians*, which is north-northwest on the dial. Not seeing π somewhere in the exponent is weird. We can make it more general, and less weird, by observing:

Proposition: Every matrix of the form $\frac{1}{2}\begin{bmatrix} 1+e^{i\theta} & 1-e^{i\theta} \\ 1-e^{i\theta} & 1+e^{i\theta} \end{bmatrix}$ is unitary.

One can show directly that the top row has dot product zero with the complex conjugate of the second row and that each row has magnitude 4, which gets multiplied by $\frac{1}{4}$ out front. Note also that if $\theta = \pi/2$, then this becomes the matrix $\mathbf{V}$ in the previous example.

For a final remark, note that A commutes with $\mathbf{I}$---well, everything commutes with the identity matrix. This exemplifies the statement in the paragraph about the "useful fact" in the second half of teh proof of the Spectral Theorem: if A and B are Hermitian matrices that commute, then they have a common orthonormal eigenbasis, but not every orthonormal eigenbasis for B is also one for A.

Speaking fairly generally, a projective measurement $\mathbf{P} = \{\mathbf{P}_i\}_{i=1}^n$ can be associated to an orthonormal eigenbasis for some Hermitian matrix. When we apply $\mathbf{P}$ to measure a general (pure) state $|\phi\rangle$, we "bonk it with the basis". The measurement outcome is one of the $|\mathbf{u}_i\rangle$ vectors, with probability $|\langle \mathbf{u}_i | \phi \rangle|^2$. (Some sources further say that the associated eigenvalue λ_i "is" the outcome.) The fact that $\mathbf{P}_i^2 = \mathbf{P}_i$ lends definiteness to the measurement result---it stays the same under "repeated bonking." *Maybe* this is what enables us to *observe* the measurement result to begin with.

(By the way, note that $B = QAQ^{-1}$ is a general representation of a change of basis transformation. But if A is notationally the identity matrix, then so is B. So the specification that $\sum_i \mathbf{P}_i = \mathbf{I}$ in the definition of projective measurement does not lock us into the *notation* for the standard basis.)

Anyway, we can give a gentle partial disagreement with the Copenhagen interpretation by saying the original quantum state doesn't "collapse"---it just gets bonked. The meaningful factor going forward is: what is the role of the choice of basis to bonk it with? And is there free will in that choice?

Trace, Density Matrices, and Measurements

When we do $|\phi\rangle\langle\phi|$ for a quantum state vector $\phi = [a_1, \ldots, a_N]^T$, the diagonal entries $a_i\overline{a_i}$ of the outerproduct give $||\phi||_2^2$, which equals 1 since ϕ is a unit vector. Since a **density matrix** ρ is a linear combination of outerproducts $|\phi\rangle\langle\phi|$ by weights summing to 1, the trace $Tr(\rho)$ is also 1. Now when a unitary matrix U acts on ϕ, the density matrix of the resulting vector $U\phi$ is

$$|U\phi\rangle\langle U\phi| \;=\; |U\phi\rangle\langle\phi|U^* \;=\; U|\phi\rangle\langle\phi|U^*.$$

By linear additivity, a unitary operator acts on a mixed state ρ by the **double action** $U\rho U^*$. Put all this together, and the rule is that *the trace of a density matrix is always* 1. The action by unitary matrices preserves the trace. Ultimately this is just the idea of probabilities summing to 1.

These ideas play into the most general idea of measurement on which there is wide consensus. It generalizes the notion of a projective measurement of a pure state. Recall that a **positive semidefinite** (**PSD**) matrix is one of the form M^*M for some matrix M.

Definition: A **positive operator valued measure** (**POVM**) is a set $\{E_1, \ldots, E_m\}$ of PSD matrices such that $\sum_{j=1}^{m} E_j \;=\; \mathbf{I}$ (text has $E_j^*E_j$ there, is it a typo?). Given a mixed state ρ the probability p_j of outcome j is given by

$$p_j \;=\; Tr(E_j\rho).$$

If a PSD representation $E_j \;=\; M_j^*M_j$ is specified for each j (it might not be unique, but specifying it is part of the measurement apparatus) then the next state is

$$\rho' \;=\; \frac{M_j\rho M_j^*}{p_j}.$$

We can use this to answer a natural question: How does the mixed state $\frac{1}{2}\big(|0\rangle + |1\rangle\big)$ differ from the quantum superposition $|+\rangle \;=\; \frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big)$? Besides the different constant, there is a difference in meaning that dictates that when mixed states are involved, we really need to use the density matrix representation of both. So we are really talking about $\frac{1}{2}\big(|0\rangle\langle 0| + |1\rangle\langle 1|\big)$ versus $|+\rangle\langle +|$. We have that the former is

$$\frac{1}{2}\left(\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\right) \;=\; \begin{bmatrix}0.5 & 0\\ 0 & 0.5\end{bmatrix}$$

while the latter is

$$\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \;=\; \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}.$$

Both matrices have trace 1; the difference is that $\mathbf{J} = |+\rangle\langle+|$ hasd non-zero off-diagonal elements. Also $\mathbf{J}^2 = \mathbf{J}$, which is the definition of when a density matrix represents a pure state. Now we know that $\mathbf{H}|+\rangle = |0\rangle$, which the density matrix under the double-action rogers:

$$\mathbf{H}\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}\mathbf{H} \;=\; \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \;=\; \frac{1}{4}\begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \;=\; \frac{1}{4}\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \;=\; |0\rangle\langle 0|,$$

whereas

$$\mathbf{H}\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\mathbf{H} \;=\; \frac{1}{2}\mathbf{H}\mathbf{H} \;=\; \frac{1}{2}\mathbf{I} \;=\; \frac{1}{2}\big(|0\rangle\langle 0| + |1\rangle\langle 1|\big)$$

back again. So whereas Alice would measure 0 with certainty if she applied Hadamard to $|+\rangle$, when she does so to her mixed state she will still get 0 with only 50% probability. The kicker is that if she instead measures in the $\{|+\rangle, |-\rangle\}$ basis, whether before or after applying the Hadamard gate, she will get those outcomes with 50% probability each. Thus "a mixed state does not remember which pure states were used to define it." The only reality it has---at least the only reality that we can *know*---is its density matrix.

[Coverage for Prelim II stops here.]

Traceout and Spectral Purification

A further rule involving density matrices and tensor products starts from pure states $|\phi\rangle$ and $|\psi\rangle$. Recall that the adjoint $(|\phi\rangle \otimes |\psi\rangle)^*$ is $\big(|\phi\rangle^* \otimes |\psi\rangle^*\big) \;=\; \langle\phi| \otimes \langle\psi|$. That is, we don't reverse the product as we would with ordinary matrix multiplication. The indexing is "tiered" in the form $state(xy)$ where x pertains to the space of $|\phi\rangle$ ("Alice") and y to the space of $|\psi\rangle$ ("Bob"). So now involving outerproducts and running indices u over Alice's row space and v over Bob's:

$$\big(|\phi\rangle \otimes |\psi\rangle\big)\big(\langle\phi| \otimes \langle\psi|\big)[uv, xy] \;=\; \big(|\phi\rangle \otimes |\psi\rangle\big)[uv]\big(\langle\phi| \otimes \langle\psi|\big)[xy] \;= \phi(u)\psi(v)\overline{\phi(x)}\,\overline{\psi(y)}.$$

Whereas,

$$\big((|\phi\rangle\langle\phi|) \otimes (|\psi\rangle\langle\psi|)\big)[uv, xy] \;=\; (|\phi\rangle\langle\phi|)[ux]\,(|\psi\rangle\langle\psi|)[vy] \;=\; \phi(u)\overline{\phi(x)}\psi(v)\overline{\psi(y)},$$

which is the same. Note that the left-hand side of the second equation is the tensor product of two pure density matrices. By additive linearity for tensor products, this proves the general rule:

> The density matrix of two unentangled systems can be represented as the tensor product of density matrices of the respective systems. In symbols: $\rho_{A,B} = \rho_A \otimes \rho_B$. (Here we understand identity up to multiplication by unit scalars.)

A nifty point is that we can semi-invert this process *even when Alice and Bob are entangled*. The operation is called the **traceout**. It is easiest to picture and execute when we apply it to the second tier of the whole space, i.e., in *"tracing out Bob."* It is also called the **partial trace** $Tr_{\mathbb{B}}$ mapping elements of the "higher space" $\mathbb{A} \otimes \mathbb{B}$ to the space $\mathbb{A}$. Given the density matrix ρ of the whole system:

- Block out ρ into square submatrices as-if it were a tensor product $A \otimes B$. If Bob holds k qubits, then the submatrices will be $2^k \times 2^k$.
- Replace each submatrix by its trace. When you consider the submatrices on the main diagonal, you can see the overall trace is unchanged---it is still 1 as it must be for ρ.
- The resulting matrix is the density ρ_A for Alice "after tracing out Bob."

There is also a matrx ρ_B of Bob "tracing out Alice." However, it need not follow that $\rho = \rho_A \otimes \rho_B$. That happens if (and only if? the things that occur to you on the second pass...) Alice and Bob were initially unentangled. In that case, *all* of the "Bob" submatrices have trace 1. The effect is the same---in the case above where Alice and Bob are pure states---as substituting $\psi(v) = \psi(y) = 1$. This leaves $\phi(u)\overline{\phi(x)}$, which is the ux entry of Alice's outerproduct $|\phi\rangle\langle\phi|$.

Example 1: The traceout of the entangled state $\frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big)$ is done by

$$Tr_B\left(\frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix},$$

which is the density matrix of the completely mixed state again. Notice incidentally that $\frac{1}{\sqrt{2}}\big(|{+}{+}\rangle + |{-}{-}\rangle\big)$ gives exactly the same pure-state vector $\frac{1}{\sqrt{2}}[1,0,0,1]^T$, even before we take its outerproduct to get the above 4×4 density matrix. This all reinforces that Alice applying $\mathbf{H}$ or whatever unitary operation on her half of the entangled pair has no effect on the current state of her knowledge of it, which is represented by the density matrix.

Example 2: The pure state $\frac{1}{2}\big(|000\rangle + |001\rangle + |110\rangle - |111\rangle\big)$ has the following density matrix:

$$\frac{1}{4}\left[\begin{array}{c|cccccccc} & 000 & 001 & 010 & 011 & 100 & 101 & 110 & -111 \\ \hline 000 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 001 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 010 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 011 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 100 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 101 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 110 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 \\ -111 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 \end{array}\right]$$

Notice that the two off-diagonal traces cancel. So tracing out the third qubit leaves:

$$\frac{1}{4}\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}.$$

This is not the density matrix of an entangled pair. Nor is it even the completely mixed state on two qubits. It is the mixture $0.5\big(|00\rangle\langle 00| + |11\rangle\langle 11|\big)$. Its tensor product with Bob holding $|+\rangle$ is not the same as the above density matrix---the corners stay zeroed out. This is another indication that our original pure state is entangled.

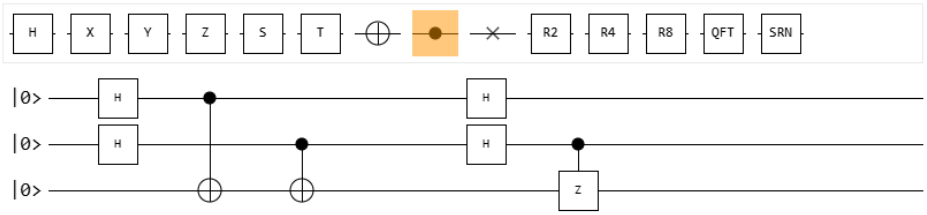
[The Thu. 11/13 lecture ended here. The Tue. 11/18 lecture will pick up here.]

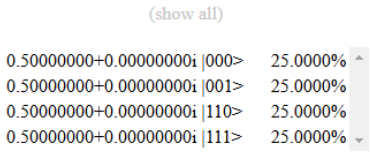
Whereas, if we use a **CCZ** gate---or just a **CZ** gate on the second and third qubits---then the minus sign flips to make $\frac{1}{2}\big(|000\rangle + |001\rangle + |110\rangle + |111\rangle\big)$. Now the pure density matrix is

$$\frac{1}{4}\left[\begin{array}{c|cccccccc} & 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ \hline 000 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 001 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 010 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 011 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 100 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 101 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 110 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 111 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right]$$

Tracing out Bob does leave the density matrix $\frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$ of the pure state $\frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big)$ in

Alice's hands. And the whole system truly is $(\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Here is the quantum circuit, including the final **CZ** gate:





The interpretation that might go with this is that the first two Hadamard and CNOT gates tried to entangle qubits 1 and 3 and then entangled 2 and 3. The hope was that by "entangling 3 twice" we could actually *disentangle* it and leave 1 and 2 entangled. This didn't quite work, but it does work if we apply the **CZ** gate after the second Hadamard gate. The subject of **entanglement swapping** commonly needs 6 qubits to illustrate, but this gives some of the flavor at smaller scale.

One more notable fact:

Theorem: For every mixed state ρ on n qubits there is a pure state $|\kappa\rangle$ on $2n$ qubits such that tracing out the last n qubits in $|\kappa\rangle\langle\kappa|$ leaves exactly ρ.

Proof: Because ρ is Hermitian, we can find an orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_N$ (where $N = 2^n$ as usual) and real eigenvalues $\lambda_1, \ldots, \lambda_N$ such that $\rho = \lambda_1|\mathbf{u}_1\rangle\langle\mathbf{u}_1| + \cdots + \lambda_N|\mathbf{u}_N\rangle\langle\mathbf{u}_N|$. This gives a diagonal matrix in the coordinates of the eigenbasis, but we can apply a unitary change of basis U to make $U\rho U^*$ diagonal in the standard basis. Since we've seen that this double-action preserves the trace, which is 1 in ρ, we get $\lambda_1 + \cdots + \lambda_N = 1$. (Note how this says ρ is far different from a unitary matrix, even a unitary matrix that is Hermitian, because those have each individual eigenvalue being of magnitude 1.) Now define

$$\kappa = \sqrt{\lambda_1}(|\mathbf{u}_1\rangle \otimes |\mathbf{u}_1\rangle) + \cdots + \sqrt{\lambda_N}(|\mathbf{u}_N\rangle \otimes |\mathbf{u}_N\rangle).$$

This is a legal pure state because the squares of the amplitudes $\sqrt{\lambda_i}$ sum to 1. So let us apply the traceout to $|\kappa\rangle\langle\kappa|$.

When we do $|\kappa\rangle\langle\kappa|$, we get cross-terms but they stay within each n-qubit tier of the whole Hilbert space---by the point we observed at the beginning of this section. Within each tier, they have the form $|\mathbf{u}_i\rangle\langle\mathbf{u}_j|$ with $i \neq j$ (multiplied by $\sqrt{\lambda_i\lambda_j}$). Now the main diagonal of this outerproduct is $\sum_{k=1}^{N} \mathbf{u}_i(k)\overline{\mathbf{u}_j(k)}$, which is exactly the *inner* product $\langle\mathbf{u}_j|\mathbf{u}_i\rangle$. This in turn is zero because $\mathbf{u}_i$ and $\mathbf{u}_j$ are orthogonal. So taking the trace of these ``Bob'' submatrices makes the off-diagonal components of the traceout vanish without a trace. The only survivors are the terms

$$\left|\sqrt{\lambda_i}\mathbf{u}_i \otimes \mathbf{u}_i\right\rangle\left\langle\sqrt{\lambda_i}\mathbf{u}_i \otimes \mathbf{u}_i\right| \;=\; \lambda_i \left|\mathbf{u}_i \otimes \mathbf{u}_i\right\rangle\left\langle\mathbf{u}_i \otimes \mathbf{u}_i\right|.$$

Now tracing out "Bob" in these submatrices just substitutes 1 for the second $\mathbf{u}_i$, leaving

$$\lambda_1 \left|\mathbf{u}_1\right\rangle\left\langle\mathbf{u}_1\right| + \;\cdots\; + \lambda_N \left|\mathbf{u}_N\right\rangle\left\langle\mathbf{u}_N\right|,$$

which is the original ρ back again. ☒

There are other possible pure states on higher numbers of qubits that can do the same. In Example 2 above, we saw that the 2-qubit mixture $0.5\left(|00\rangle\langle 00| + |11\rangle\langle 11|\right)$ is the traceout of the 3-qubit pure state $\frac{1}{2}\left(|000\rangle + |001\rangle + |110\rangle - |111\rangle\right)$, where we got the off-diagonal cancellations without needing to go to a full-blown spectral representation. Moreover, the Bell pair $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$ is *exactly* the $|\kappa\rangle$ from the completely mixed state $\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$, which is already in spectral form. The general name for this process is **mixed-state purification**. It often happens that the neatest way to calculate or prove results about mixed states is to "lift" them to pure states in a larger space, calculate in the higher space, and then trace back down. John Smolin of IBM T.J. Watson gave this technique the evocative name of "appealing to the church of the higher Hilbert Space."

Choosing Bases to Measure In

The question that concerned Einstein is whether Bob can send a willful message to Alice through their entanglement by choices of measurement bases. My use of "willful" here is willful: *pace* quantum-based arguments against free will, it is IMHO the clearest way to frame the technical argument. All agree that Alice gains *information* of Bob's random outcomes, though that information was "pre-paid" by the interactions that set up n entangled qubits to begin with. The point of superdense coding is that Bob could distinguish among 4 willful actions by Alice *after* the initial exchange of one entnagled qubit, when it was *followed by* her sending 1 other qubit. Can something like this be done *without any further interaction*---and over time intervals shorter than the time for light to travel between Alice and Bob?

Most in particular, can Alice gain any willful information---other than unstructured randomness---from how Bob orients his measurements? The answer is **no**. If they share $|00\rangle + |11\rangle$ (over $\sqrt{2}$) you might think Bob could guarantee a '1' by measuring in the $|+\rangle, |-\rangle$ basis, but no: that was the first decoherence example with Alice. Any basis Bob uses is the same as a unitary $\mathbf{U}$ to convert to the standard basis followed by a measurement there, and $\mathbf{U}$ has no effect on what Alice will see.

This makes it all the more amazing that there are situations where the choice of measurement basis does make a difference---one that has been quantified in actual experiments. This comes next.

The CHSH Game

The initials in the CHSH Game stand for John Clauser, Michael Horne, Abner Shimony, and Richard A. Holt, who described it in a paper in 1969. The 2022 Nobel Prize in Physics was awarded to Clauser and to Alain Aspect and Anton Zeilinger. The latter two experimentally confirmed the quantum advantage.

In the game, Alice and Bob share n Bell pairs $\frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)$ and can have as much prior classical communication to agree on strategies as they please. Between the start and end of a *trial*---one play of the game---they may not communicate with each other, but they may observe common sources. The common source can not only be random---such as from patterns of solar flares both Alice and Bob can see---it can be controlled by an oracle "Ozzie" who is trying to help Alice and Bob. Each trial operates via classical communication with a third party, "Ralph" (to sound like ref, referee) and goes like this:

1. Ralph sends a random bit a to Alice and a bit b to Bob. Neither can see the other's bit.
2. Alice sends a response bit u to Ralph and Bob simultaneously sends his response bit v to Ralph.
3. Ralph declares that Alice and Bob win the trial if $u \oplus v = a \wedge b$.

We may suppose that Alice and Bob receive a and b in sealed boxes, and give their respective u and v within a nanosecond of opening their boxes. Without loss of generality, we may suppose that any other influence from observations or "Ozzie" has been registered by that instant. At that point, Alice's u is a one-bit Boolean function of a alone. We use $0, 1$ for the inputs to this function but give the outputs as **Y** for "yes" or **N** for "no" in order to keep inputs and outputs visually separate. There are just four functions that she can use:

- The always-true function: yes to 0 and yes to 1, which we call **YY**.
- The always-false function, which we similarly call **NN**.
- The identity function, giving Ralph the same bit back, which is **NY**.
- Flipping the bit to Ralph, which is **YN**.

Bob has the same four options, so there are in total **16** different strategies they can use for any trial. Meanwhile, Ralph has his own four possible actions. Here is the entire matrix of possibilities. The matrix entries are numeric rather than Boolean: 1 if Alice and Bob win, 0 if they lose. The rows are the four options by Ralph, in order a, b so that for instance, if Alice and Bob adopt the strategy in the third column and find that Ralph chose $1, 0$, then Alice says $\mathbf{N}$ while Bob says $\mathbf{Y}$---and they lose because their answers disagreed while $1 \wedge 0$ is false.

Alice	NN	NN	NN	NN	NY	NY	NY	NY	YN	YN	YN	YN	YY	YY	YY	YY
Bob	NN	NY	YN	YY	NN	NY	YN	YY	NN	NY	YN	YY	NN	NY	YN	YY
$0, 0$	1	1	0	0	1	1	0	0	0	0	1	1	0	0	1	1
$0, 1$	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1
$1, 0$	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1
$1, 1$	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0

Note that when Ralph plays randomly, Alice and Bob can assure 75% winning if they choose any of the eight columns with three 1s as their joint strategy. **They cannot do better**, because every column has a case where Ralph could send them something that makes their joint strategy lose---and the randomized Ralph does so with 25% probability.

The amazing fact is that sharing one entangled Bell pair enables Alice and Bob to do much better: to win **over 85%** of the time in theory.

A side note if you are familiar with matrix game theory: could Ralph do better with non-random play if he knew Alice and Bob's strategy? Certainly if Alice and Bob always play one fixed column, such as both always saying N, then Ralph could always deny them by giving the one losing combination a, b. If you know the **Minimax Theorem** of zero-sum matrix game theory, then you already know that because 25% is Ralph's optimum when he has to move first, Alice and Bob *must* have a *classically randomized* strategy that assures them 75% even if Ralph is told about it in advance. We can find it easily by first removing the eight "obviously stupid" joint strategies---those with only one 1 in their column---leaving:

Alice	NN	NN	NY	NY	YN	YN	YY	YY
Bob	NN	NY	NN	YN	NY	YY	YN	YY
0, 0	1	1	1	0	0	1	1	1
0, 1	1	0	1	1	1	1	0	1
1, 0	1	1	0	1	1	0	1	1
1, 1	0	1	1	1	1	1	1	0

Now if Alice and Bob use their shared classical randomness to choose one of the leftover strategies at random with probability $1/8$, there is no way Ralph can avoid their winning 75% of the time even if Ralph knows that is their policy. If Ralph could steal their random bits by looking at solar flares *and* knowing *how* and *when* Alice and Bob will decode them, then Ralph could still always send the bad combo. But the order is: Ralph commits to the a, b combo first, then Alice and Bob have a moment to read the shared random source that determines their policies before they open their boxes. The scientific significance does not require this detail---we just stipulate that Ralph plays randomly.

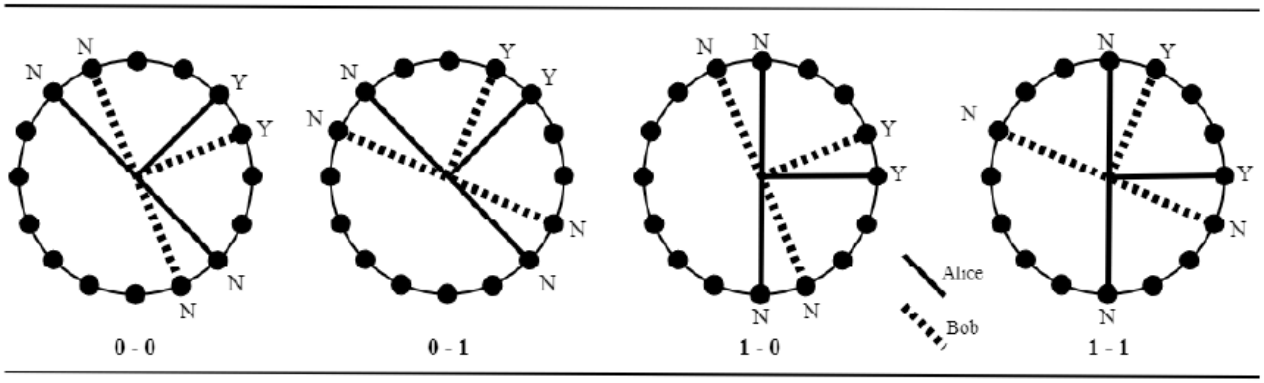
The Quantum Case

Alice and Bob get an extra option using one shared Bell pair per trial: Each can measure in a basis that depends on the bit received from Ralph. The timing of this option is synchronized as viewed by Ralph. The text describes Alice as measuring first, but we'll make Bob go first for consistency with recent lectures. By symmetry, it does not matter who goes first. What does matter, technically, is that the time lapse from opening the boxes to the second measurement---as viewed by Ralph---must be less than the time it would take light to travel from Bob to Alice. This is in order to avoid one of several possible "loopholes" that could enable a classical explanation.

Rather than the Bloch sphere, this is a case where the Cartesian diagram of state vectors is best for visualization: $|0\rangle$ at east (**E**), $|1\rangle$ at north (**N**), $|+\rangle$ between them facing northeast (**NE**), and $|-\rangle$ to the southeast (**SE**). Alice will use either the $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ or $\{|+\rangle\langle +|, |-\rangle\langle -|\}$ measurement. We let the former outcomes stand for "yes", so we can abbreviate her options as **E** or **NE**. Bob has a funkier set of bases to choose from. He can use the basis that orients his "yes" answer at 22.5°, which we call **ENE** for east-northeast, and puts "no" at 112.5° (or equivalently, at -67.5°, i.e., 292.5°). Or Bob can use the basis that puts "yes" at $+67.5^\circ$, which is **NNE** for north-northeast. Here is the protocol:

1. Alice and Bob open their boxes simultaneously.
2. If $b = 0$, Bob measures his entangled qubit in the basis oriented **ENE**; if $b = 1$, Bob chooses **NNE**.
3. If $a = 0$, Alice instantly chooses **NE**, else she chooses **E**. No more than a nanosecond later than Bob's actions, Alice measures her qubit in her chosen basis.
4. Each sends Ralph "yes" if getting the measurement outcome designated "yes", else "no".

The upshot can be appreciated *ahead of any thinking about the underlying physical reality*, just by looking at the diagram of the choices made by Alice and Bob in the four cases Ralph can send them:



Alice's chosen orientations depend only on her bit from Ralph: northeast on 0, due east on 1. Bob likewise reacts independently of Alice. Yet the options combine to make their "yes" orientations come within 22.5° of each other in all of the 00, 01, and 10 cases from Ralph, yet 67.5° apart on 11.

If being one-fourth of a right angle apart meant a one-fourth chance of losing, then the resulting chances would be no different from the classical case: 75% frequency of winning. But in ways we can actually see for ourselves by orienting polarizing filters at these angles and telling how much light gets through, in the first three cases, the chance of her qubit instantly transformed(?) by Bob's outcome giving the same yes/no answer from her 22.5°-apart measurement is greater: $\cos^2\left(\frac{\pi}{8}\right) = 85.3553...\,\%$.

And in the fourth case, the frequency of Alice and Bob giving different answers and winning is the same.

Discussion

Well, saying "transformed" is exactly the kind of *spukhafte Fernwirkung* that Einstein objected to. But this is the straightest path to expressing the explanation for what we observe---which has been verified in actual experiments achieving over 80%. The gap between 80% and 85+% is ascribable in substantial part to the kind of slight-degrading errors we saw in the "depolarization and de-phasing" section (plus to other slight flaws in the apparatus and its nanosecond timing).

Note that no "free will" is involved on the part of Alice and Bob, nor any contextual information ("hints from Ozzie") at all. Their choices of measurement basis are determined entirely by the bit each receives from Ralph. Their only agency is the sharing of entangled Bell pairs, possession of the measuring apparatus for their respective pairs of bases, and a mechanism for reading the bit from Ralph and effecting the corresponding basis choice. Given a physical setup and timing so that their measurements are made within a picosecond of receiving the bit from Ralph and of each other, while "Alice" and "Bob" are situated more than a light-picosecond apart, Ralph is really playing solitaire. And Ralph plays randomly, so no free will is involved there either. Yet the resulting physical system "wins" with a frequency that cannot be explained by any classical theory with variables localized to "Alice" and "Bob" that obliviates the entanglements between them.

My section 14.7.3 replaces the element of Alice and Bob choosing different measurement bases with that of their choosing different *basis-change operators*, while always doing their actual measurements in the standard basis. They apply these operators **before** (and only nominally after) doing their measurements. This streamlines the physical interpretation, and yet yields the same basic math. See also the chapter end notes for further discussion. This should go hand-in-hand with the No-Communication Theorem, but the Wikipedia treatment which I've linked goes a little further afield than I had in mind for the textbook.

Finally, this example avoids objections to earlier claims of "quantum advantage"---by which the Deutsch and Deutsch-Jozsa algorithms "unfairly" restrict the classical setting; Simon's algorithm is has a discrepancy between quantum and classical that is provable but only asymptotic; Shor's algorithm is proven but factoring might be in classical (random) polynomial time after all; and Grover's algorithm gives speedups only for running times that are exponential to begin with. The ability to win more than the classical limit of 75% is concrete and experimentally proven. The only knock is that the CHSH game is for an interactive protocol, not for straight-up computation.

Section 14.8 gives a claim of quantum advantage for straight-up computation, but it has come under more of a cloud since its October 2019 unveiling (see this article by me), and is for a contrived problem anyway. We will instead seque into ideas for classical computing to take away the appearance of quantum advantage for straight-up computational problems.

Quantum Ontology and Epistemology

Ontology has to do with *being*; *epistemology* with *knowing*. We have taken the view that pure states "are"---that is, they have existence unto themselves. We represent them as state vectors, but at exponentially high cost in many cases. For properly mixed states ρ, this is less clear. We can regard some pure state $|\kappa\rangle$ from a higher space that traces out to it as its ontology, but (a) $|\kappa\rangle$ is far from unique, and (b) as indicated by the use of "N" in the proof of it, it often comes at exponential cost.

The epistemological side, however, has given a remarkably consistent set of answers for over a century now:

- The only way we can gain knowledge about a quantum state, whether pure or mixed, is by measuring it.
- All measurements of a pure state $|\phi\rangle$ go---explicitly or implicitly---through its density matrix $\rho_\phi = |\phi\rangle\langle\phi|$.
- Operations on density matrices "gibe" with measurements and probabilities in ways already prescribed by (Bayesian!) conditional inferencing.
- All scalar quantities involved in this reckoning are *real* numbers denoting (conditional) *probabilities*, not "amplitudes".

The last point is part of why Richard Lipton and I have mused about giving an account of quantum reality without complex numbers. For the above, where Hermitian not unitary matrices are primary, complex numbers need only be seen in components of orthonormal eigenvectors $\mathbf{u}_i$, such as $\mathbf{u}_1 = \frac{1}{\sqrt{2}}[1, i]^T$ and $\mathbf{u}_2 = \frac{1}{\sqrt{2}}[i, 1]^T$ on both the practice and actual Prelim II. For reality, however, the notion of *phase* seems inescapable, and complex numbers (IMPHO) give its best treatment. There are wide indications that Bloch spheres---in higher dimensions as well as for single qubits---are physically real. They give a description via two real numbers θ and φ; note that $\cos(\theta)$ is a probability, not an amplitude. However, φ is a phase angle and governs whether and where complex numbers appear in other figuring. So *Reali*ty strikes back but doesn't completely subjugate the complex realm, which is necessitated by the Fundamental Theorem of Algebra anyway.

This finally leaves the super-skeptical question of whether there is a bedrock of *being* beneath what is knowable. One form of this question is whether the notion of an *observer*---often styled as a *conscious* observer---is essential to existence. This idea long predates quantum mechanics. It was formulated as philosophical *subjective idealism* by the Irish Anglican bishop George Berkeley in the early 1700s, whom the city of Berkeley in California is expressly named after. It is well captured by the following pair of limericks---my mod of what Fr. Ronald Knox wrote two centuries later:

A divinity student said, "God
Must find it exceedingly odd
That the Warden's plum tree
Continues to be
When there's no one about in the Quad."

And the reply as Knox imagined in a newspaper's Letters column:

> "Dear Sir, your perplexity's odd.
> *I* am always about in the Quad.
> So the Warden's plum tree
> Shall continue to be,
> Since observed by---
> yours faithfully, ---God."

Whether the advent of quantum mechanics enhances such arguments over the pervasive presence of an unseen benevolent God is not something I choose to amplify. Lipton and I broadly sympathize with Samuel Johnson's reply of refuting Berkeley by kicking a stone---fully aware that the kick involves the exertion of quantum mechanical electric force on a surface whose solidity is effected by vibrating molecules. Speaking for myself as a Christian, I hold a halfway position toward *fideism* that disclaims logical proof and reproducible knowledge of God, and I regard this as merely orthodox.

There is, however, considerable reason to assert the pervasive presence of an unseen---and vaguely malign(?)---"Bob" in the form of entanglements with unknown systems, even going back to the Big Bang. Entanglements with outside nature, developed both now and prior, are the current best explanation for *decoherence*. The above illustrations of Tom Brady-style "deflation" in off-diagonal parts of density matrices are well representative of decoherence.

Staying completely with "Nature's Rose", we will conclude with the matter of classical simulation of quantum algorithms, via advanced computational methods. First on the list is the Singular Value Decomposition, which is the closest an arbitrary---not even square---matrix can come to the blessings of both unitary and Hermitian properties.