Physics 130C Lecture Notes
Chapter 1: Quantum Mechanics in Your Face*

Lecturer: McGreevy

Last updated: 2013/03/01, 15:21:35

*The title of this chapter is a reference to this lecture by Sidney Coleman, which gives a brilliant summary of the first part of our course. These lectures borrow from many sources, including: Preskill's notes, Eugene Commins lectures at Berkeley, lectures by M. Kreuzer, the book by Schumacher and Westmoreland, the book by Rieffel and Polak.

1.1 Introductory remarks

I want to say some words about my goals for this course. Probably they are too ambitious; we will see.

My first goal is to try to distill some of the strangeness inherent to quantum mechanics from the complication involved in studying infinite-dimensional Hilbert spaces (and the associated issues of functional analysis, Fourier transforms, differential equations...). Grappling with the latter issues is, no question, very important for becoming a practicing quantum mechanic. However, these difficulties can obscure (a) the basic simplicity of quantum mechanics and (b) its fundamental weirdness. With this in mind we are going to spend a lot of time with discrete systems, and finite-dimensional Hilbert spaces, and we are going to think hard about what it means to live in a quantum world. In particular, there are things that quantum computers can do faster than classical computers. There are even some things that simply can't be done classically that can be done quantumly.

A second specific goal follows naturally from this; it is sort of the converse. I want to see to what extent we can understand the following thing. The world really seems classical to us, in the sense that despite the fact that we all by now believe that the world and everything that goes on it is described by quantum mechanics, and we all experience the world and things going on in it all the time, quantum mechanics seems unfamiliar to us. How can this be? Why is it that a quantum mechanical world seems classical to its conscious inhabitants?

This is a scientific question! And it's one that hasn't been completely solved, but a lot of progress has been made.

Finally, it is worth reminding ourselves of some basic mysteries of nature – obvious facts that don't require lots of fancy technology to investigate – which are answered by quantum mechanics. It is easy to get bogged down by the complications of learning the subject and forget about these things. For example:

- *The "UV catastrophe"*: How much energy is in a box of empty space when we heat it up? How does it manage to be finite?
- *Stability of atoms:* We know from Rutherford that an atom has a small nucleus of positive charge surrounded by electrons somehow in orbit around it. Moving in a circular orbit means accelerating toward the center of the circle; accelerating a charged particle causes it to radiate. So why are atoms stable?
- Why does the sun shine? Why is it yellow?
- Why are metals such good conductors? We know how far apart the atoms are in a chunk of copper. Think about a (classical cartoon of an) electron getting pushed along by an external electric field and ask what resistivity we get using this distance as the mean free path. The answer is too large by many orders of magnitude.

- What is going on with the heat capacity of solids?
- What can happen when we let electrons hop around in a crystal and interact strongly with each other?

We are not going to answer all of these questions in detail, but we will touch on many of them. In case you can't tell already from these examples, I am going to try to make contact whenever I can with entry points to fields of current physics research (with a big bias towards theoretical physics, since that's what I do): quantum ("hard") condensed matter physics, quantum information theory, high-energy particle physics and quantum field theory.

Finally, some requests: I have assayed a good-faith effort to find out what quantum mechanics you have learned so far. But if you see that I am about to go on at length about something that you learned well in 130A and/or 130B, please tell me at your earliest convenience. Conversely, if I am saying something for which you feel wildly unprepared, also please tell me. Also: please do peruse the list of topics in the tentative course outline, and do tell me which topics you are most interested to learn about. I am depending on your feedback to maximize the usefulness of this course.

1.2 Basic facts of quantum mechanics

The purpose of this first part of the course is to agree upon the structure of quantum mechanics[1] and upon Dirac's wonderful notation for it.

Since you have already spent two quarters grappling with quantum phenomena, I will jump right into the axioms and will not preface them with examples of the phenomena which motivate them. I'm going to state them in terms of four parts (**States, Observables, Time Evolution, Measurement**), which are, very briefly, the answers to the following questions:

1. **States:** How do we represent our knowledge of a physical system?
2. **Observables:** What can we, in principle, measure about it?
3. **Time Evolution:** If we know the state now, what happens to it later?
4. **Measurement:** What happens when we actually make a measurement?

We will discuss the first two axioms now; then we will have a lightning review of the essential linear algebra. Then we'll talk about the other two.

1. **States**

   By 'state' here, I mean a complete description of what's going on in a physical system, a specification of as much data about it as is physically possible. For example, in classical mechanics of a bunch of particles, this is a specification of the coordinates and momenta of every particle. In quantum mechanics, the state of a system is a **ray** in the **Hilbert space** of the system. This is a very strong statement about the nature of quantum mechanics whose meaning we will spend a lot of time unpacking. First I must define the terms in bold:

   For our purposes, **Hilbert space** is a vector space over the complex numbers $\mathbb{C}$, with two more properties (below). I am going to use Dirac notation to denote vectors, which looks like this: $|\psi\rangle$[2]. We can add vectors to get another vector: $|a\rangle + |b\rangle$; we can multiply a vector by a complex number to get another vector: $z|a\rangle$. Please don't let me see you try to add a vector to a number, it will make me sad.

   The two other properties of a Hilbert space (beyond just being a vector space) are:

[1] [Here we follow the discussion of Preskill, Chapter 2.]

[2] Some of the beauty of this becomes clear when we notice that basis elements are associated with physical properties of the system and we can put that information in the little box (called a 'ket'). For example, if we are talking about a spin which can be up or down, there will be a state $|\uparrow\rangle$ and a state $|\downarrow\rangle$. I leave it as a fun exercise for the reader to imagine what else we might put in the little box. I will nevertheless often succumb to the silly convention of calling states things like $|\psi\rangle$ and $|a\rangle$.

1) It has an inner product. This means that given two vectors $|a\rangle$ and $|b\rangle$ I can make a complex number out of them, which we denote: $\langle a|b\rangle$ [3]. This inner product must have the following three nice features:

(a) Positivity: $\langle a|a\rangle > 0$ for $|a\rangle \neq 0$.

(b) Linearity: $\langle a|\left(z|b\rangle + w|c\rangle\right) = z\langle a|b\rangle + w\langle a|c\rangle$.

(c) Skew symmetry: $\langle a|b\rangle = \langle b|a\rangle^\star$.

Note that (c) and (b) imply that the inner product is *anti*linear in the bra vector, that is: if I ask what is the inner product between $|d\rangle \equiv (z|a\rangle + w|b\rangle)$ and $|c\rangle$ the answer is :

$$\langle d|c\rangle = z^\star\langle a|c\rangle + w^\star\langle b|c\rangle.$$

An interjection to motivate the awesomeness of the Dirac notation [Feynman III-8-1]: Consider the inner product of two vectors, which we denote by

$$\langle\chi|\phi\rangle \ .$$

For vectors in 3-space, we are used to computing this using Pythagoras, that is, by projecting the two vectors onto a preferred orthonormal basis, and adding the components[4]. In various notations, this looks like:

$$\langle\chi|\phi\rangle = \sum_i (\chi\cdot e_i)\,(e_i\cdot\phi) = \sum_i \chi_i^\star\phi_i = \chi_x^\star\phi_x + \chi_y^\star\phi_y + ... = \sum_i \langle\chi|i\rangle\langle i|\phi\rangle.$$

BUT: this is true for any χ, so we may as well erase the χ:

$$|\phi\rangle = \sum_i |i\rangle\langle i|\phi\rangle$$

and we are forced directly to Dirac's notation.

There is one further logical step we can take here. This relation is also true for all ϕ, so we can erase the ϕ, too! :

$$| = \sum_i |i\rangle\langle i| \ .$$

The little line | stands for the identity operator. I will usually write it a little fancier, as $\mathbb{1}$.

[End of Lecture 1]

[3] Note that the element of the dual space, $\langle a|$, is known as a *bra*; together the bra and the ket join (like Voltron) to form a bracket. This is a *pun* made by Dirac (!). Let it not be said that physicists don't have a sense of humor. Note also that this notation is nicely consistent with our notation for expectation values.

[4] I guess I am imagine a version of Pythagoras who knew about complex numbers.

2) The Hilbert space is *complete* in the norm determined by the inner product $\| a \| \equiv \langle a|a\rangle^{1/2}$. This means that we can form a resolution of the identity of the form

$$\mathbb{1} = \sum_a |a\rangle\langle a|.$$

For discrete, finite systems, this is obvious; For infinite-dimensional Hilbert spaces it is important[5].

I still owe you a definition of **ray**. A ray is an equivalence class of vectors, under the equivalence relation $|a\rangle \simeq z|a\rangle$. This just means that we don't distinguish between a vector and its product with a (nonzero) complex number. We can remove (some of) this ambiguity by demanding that our states be *normalized* vectors:

$$\langle a|a\rangle = 1.$$

But this doesn't stop us from changing $|a\rangle \to e^{i\varphi}|a\rangle$ for some real φ; these describe the same physical state.

Important point: it's only the *overall* phase that doesn't matter. Given two states $|a\rangle, |b\rangle$, another allowed state is the *superposition* of the two: $z|a\rangle + w|b\rangle$. The relative phase between z and w here is meaningful physical information (we will interpret it at length later); we identify

$$z|a\rangle + w|b\rangle \simeq e^{i\varphi}\left(z|a\rangle + w|b\rangle\right) \quad \neq \quad z|a\rangle + e^{i\varphi}w|b\rangle \ .$$

2. **Observables**

By observables, we mean properties of a physical system that can in principle be measured. In quantum mechanics, an observable is a **self-adjoint** (or Hermitian) **operator**.

Again with the bold symbols: An **operator** is a linear map which eats vectors and spits out vectors:

$$\mathbf{A} : |a\rangle \to \mathbf{A}|a\rangle = |\mathbf{A}a\rangle, \qquad \mathbf{A}\left(z|a\rangle + w|b\rangle\right) = z\mathbf{A}|a\rangle + w\mathbf{A}|b\rangle \ .$$

To indicate operators, I'll use boldface or put little hats, like $\hat{A}$.

The **adjoint** $\mathbf{A}^\dagger$ of an operator $\mathbf{A}$ is defined in terms of the inner product as follows:

$$\langle \mathbf{A}^\dagger a|b\rangle := \langle a|\mathbf{A}|b\rangle = \langle a|\mathbf{A}b\rangle$$

for any a, b; this is is a definition of the thing on the left; in the third expression here, I've written it in a way to emphasize that the $\mathbf{A}$ is acting to the right on b. An equivalent definition of the adjoint of an operator is: if $|v\rangle = \mathbf{A}|u\rangle$, then $\langle v| = \langle u|\mathbf{A}^\dagger$.

[5]For completeness of the discussion of completeness, here's the complete definition of *complete*: it means that any *Cauchy sequence* of vectors has a limit in the Hilbert space. A Cauchy sequence $\{v_i\}$ is one where the successive elements get close together – this is where the norm comes in: $\forall \epsilon, \exists n$ such that $\| v_i - v_j \| < \epsilon$ when $i, j > n$. This is just the kind of thing I don't want to get hung up on right now. Hence, footnote.

If we choose a basis for our Hilbert space, $\mathcal{H} = \text{span}\{|n\rangle\}$ (for n in some suitable set)[6] we can represent each operator on it as a matrix:

$$\langle m|\mathbf{A}|n\rangle \equiv A_{mn} \ .$$

In an orthonormal (ON) basis, we can write the matrix in terms of its matrix elements as

$$\mathbf{A} = \sum_{m,n} A_{mn}|m\rangle\langle n|.$$

Notice that a matrix which is *diagonal* in the basis looks like $A_{mn} = \delta_{mn}A_n$:

$$\mathbf{A} = \sum_{m,n} \delta_{m,n}A_n|m\rangle\langle n| = \sum_n A_n|n\rangle\langle n| \ ;$$

its eigenvectors are $|n\rangle$ and its eigenvalues are A_n.

Let me elaborate a little more on translating between Dirac notation and matrices. Suppose I have a Hilbert space with two dimensions, $\mathcal{H} = \text{span}\{|1\rangle, |2\rangle\}$. where $|1\rangle, |2\rangle$ are an ON basis. Let me represent vectors by their components in this basis, so :

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

In order to reproduce the ON property $\langle n|m\rangle = \delta_{nm}$, we must have

$$\langle 1| = (1,0) \qquad \text{and} \qquad |2\rangle = (0,1).$$

Then notice that I can resolve the identity in this basis as

$$\sum_n |n\rangle\langle n| = |1\rangle\langle 1| + |2\rangle\langle 2| = \begin{pmatrix} 1 \\ 0 \end{pmatrix}(1,0) + \begin{pmatrix} 0 \\ 1 \end{pmatrix}(0,1) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ .$$

By the way, this kind of operation which takes two vectors and makes a (rank-one) matrix by

$$|v\rangle\langle w| = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \end{pmatrix}(w_1, w_2, \cdots) = \begin{pmatrix} v_1w_1 & v_1w_2 & \cdots \\ v_2w_1 & v_2w_2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

is called an *outer* product (or Kronecker product); it is kind of the opposite of the inner product which makes a scalar $\langle v|w\rangle$.

The matrix associated with the adjoint of an operator is the conjugate transpose of the matrix representation of the operator:

$$\langle m|\mathbf{A}^\dagger|n\rangle = A^\star_{nm} = (A^{\star t})_{mn}.$$

[6] This notation "span" means: span$\{|n\rangle\}$ is the vector space whose elements are arbitrary linear combinations of the $|n\rangle$s.

Transpose means we reverse the indices: $(A^t)_{mn} \equiv A_{nm}$: this flips the elements across the diagonal, *e.g.*:

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

Sensibly, an operator is **self-adjoint** if $\mathbf{A} = \mathbf{A}^\dagger$.

1.3 Linear algebra essentials for QM

This section is a mathematical interlude. No physics will intrude on our discussion. [See Chapter 2 of Le Bellac, Chapter 1 of Shankar for more reading, *e.g.* if you don't believe me about something and want to appeal to authority.] Note that I am going to assume below that we are talking about a finite-dimensional Hilbert space; this avoids the annoying complications about convergence of integrals.

A **Hermitian** operator $\mathbf{A}$ is equal to its adjoint $\mathbf{A}^\dagger = \mathbf{A}$. [7]

- Hermitian operators have real eigenvalues and orthonormal (ON) eigenbases.

Proof: Let $|\alpha\rangle, |\beta\rangle$ be normalized eigenvectors of $\mathbf{A}$. So:

$$\mathbf{A}|\alpha\rangle = \alpha|\alpha\rangle \tag{1}$$

$$\text{and} \qquad \mathbf{A}|\beta\rangle = \beta|\beta\rangle \implies \langle\beta|\mathbf{A}^\dagger = \langle\beta|\beta^\star \ .$$

Hit (1) on the left with $\langle\beta|$:

$$\langle\beta|\mathbf{A}|\alpha\rangle = \alpha\langle\beta|\alpha\rangle = \beta^\star\langle\beta|\alpha\rangle \ .$$

This implies

$$0 = (\beta^\star - \alpha)\,\langle\beta|\alpha\rangle.$$

So: if $\alpha = \beta$, $\langle\alpha|\alpha\rangle = 1$ so we learn that $\alpha = \alpha^\star$: the eigenvalues are real.

If $\alpha \neq \beta$, and the eigenvalues are distinct we learn that $\langle\beta|\alpha\rangle = 0$, the eigenvectors are orthogonal. (They are not automatically normalized, since $a|\alpha\rangle$ is also an eigenvector with the same eigenvalue. It is up to us to normalize them (for example when Mathematica spits them out.))

If the eigenvalues are degenerate (different vectors give the same eigenvalue), it is up to us to choose an orthonormal basis for the degenerate eigenspace (for example by finding a complete set of commuting operators, see below).

[7] For finite-dimensional $\mathcal{H}$, self-adjoint and Hermitian are the same. In the infinite-dimensional case, a Hermitian operator is self-adjoint and *bounded*, which means $\| \mathbf{A}|v\rangle \| < c \| |v\rangle \|, \forall v$ and for some constant c.

Note that a real symmetric matrix is a special case of a Hermitian matrix.

• Spectral decomposition: in terms of such an ON basis of eigenvectors of a Hermitian operator, we can make a super-useful resolution of the identity: If

$$\mathbf{A}|\alpha\rangle = \alpha|\alpha\rangle$$

then the identity operator, the one which does nothing to everybody, is:

$$\mathbb{1} = \sum_{\text{all } \alpha} |\alpha\rangle\langle\alpha| = \sum_{\alpha} \mathbf{P}_\alpha.$$

The object $\mathbf{P}_\alpha \equiv |\alpha\rangle\langle\alpha|$ is a *projector* onto the eigenstate $|\alpha\rangle$: $\mathbf{P}_\alpha^2 = \mathbf{P}_\alpha$. Notice that in this basis

$$\mathbf{A} = \sum_{\text{all } \alpha} \alpha|\alpha\rangle\langle\alpha|.$$

This is what a diagonal matrix looks like in Dirac notation. Note that this really does depend on Hermiticity of $\mathbf{A}$; the eigenvectors of a general matrix are not orthogonal and the sum of their projectors will not give $\mathbb{1}$ (try it in Mathematica! or if you are feeling lazy or are unfamiliar with Mathematica you could look at the notebook where I did it. It's here.).

Actually, the last paragraph was true as long as the eigenvalue α is non-degenerate; if $\mathbf{A}$ has degenerate eigenvalues, it is a problem for our scheme of labeling the states by eigenvalues of $\mathbf{A}$. We can still write

$$\mathbf{A} = \sum_{\text{all } \alpha} \alpha \mathbf{P}_\alpha$$

where now $\mathbf{P}_\alpha$ to be understood as the projector onto the (usually one-dimensional) space spanned by the eigenvectors with eigenvalue α. These operators satisfy

$$\mathbf{P}_n\mathbf{P}_m = \delta_{nm}\mathbf{P}_m, \qquad \mathbf{P}_m^\dagger = \mathbf{P}_m \quad .$$

This is again the statement that eigenvectors of a Hermitian operator associated with different eignevalues are orthogonal.

• Operators which commute $\mathbf{AB} = \mathbf{BA}$ can be simultaneously diagonalized: *i.e.* we can find a basis in which they are both diagonal. (We will denote the *commutator* of two operators by $[\mathbf{A}, \mathbf{B}] \equiv \mathbf{AB} - \mathbf{BA}$.)

Idea of proof: Consider an eigenvector of $\mathbf{A}$, $\mathbf{A}|a\rangle = a|a\rangle$. If $[\mathbf{A}, \mathbf{B}] = 0$ then we have

$$\mathbf{A}\left(\mathbf{B}|a\rangle\right) = \mathbf{B}\left(\mathbf{A}|a\rangle\right) = \mathbf{B}\left(a|a\rangle\right) = a\left(\mathbf{B}|a\rangle\right)$$

(the parentheses are just to direct your attention). This equation says that $\mathbf{B}|a\rangle$ is ALSO an eigenvector of $\mathbf{A}$ with eigenvalue a. SO: by the theorem above, if the eigenvalue a is a

non-degenerate eigenvalue of $\mathbf{A}$, then we learn that this vector must also point along $|a\rangle$: $\mathbf{B}|a\rangle \propto |a\rangle$ that is

$$\mathbf{B}|a\rangle = b|a\rangle$$

for some complex number b, which we see is an eigenvalue of b.

If there is a degenerate eigenspace of $\mathbf{A}$ with eigenvalue a (with dimension > 1), then the action of $\mathbf{B}$ generates another element of the subspace. That is: $\mathbf{B}|a\rangle$ is not necessarily *parallel* to $|a\rangle$. (Which doesn't mean that it is orthogonal to $|a\rangle$.) It generates another vector in the subspace of eigenvectors of $\mathbf{A}$ of eigenvalue a. We can then diagonalize $\mathbf{B}$ within this subspace and label a nice orthonormal basis for the subspace as $|a, b\rangle$, by the eigenvalue of $\mathbf{A}$ and those of $\mathbf{B}$. If there is still a degeneracy, you need to find another operator.

This leads us to the useful notion of a *complete set of commuting operators*. A complete set of commuting operators allow us to specify an orthonormal basis of $\mathcal{H}$ by their eigenvalues. If we have in our hands an operator with a completely non-degenerate spectrum (no two eigenvalues are equael), then it is a complete set by itself. For example: spectrum of the position operator $\hat{x}$ for a particle on a line provides a nice ON basis for that Hilbert space (as does the momentum operator).

- A *unitary* operator is one which preserves the norms of states:

$$\| \hat{U}|\psi\rangle \|^2 = \| \, |\psi\rangle \, \|^2 \quad \forall \, |\psi\rangle.$$

This means that

$$\hat{U}\hat{U}^\dagger = 1, \qquad \text{and} \qquad \hat{U}^\dagger = \hat{U}^{-1}.$$

Besides their role in time evolution (Axiom 3 below), unitary operators (aka unitary transformations) are important for the following reason, which explains why they are sometimes called 'transformations': they implement changes of basis.

To see that a basis change is implemented by a unitary operator, suppose we are given two ON bases for our $\mathcal{H}$: $\{|n\rangle, n = 1..N\}$ and $\{|a_n\rangle, n = 1..N\}$, and a 1-to-1 correspondence between the two $|n\rangle \leftrightarrow |a_n\rangle$. *Define* $\mathbf{U}$ to be the linear operation which takes $\mathbf{U}|n\rangle = |a_n\rangle$. Taking the adjoint gives $\langle n|\mathbf{U}^\dagger = \langle a_n|$. Then

$$\mathbf{U} = \mathbf{U}\mathbb{1} = \mathbf{U}\sum_n |n\rangle\langle n| = \sum_n (\mathbf{U}|n\rangle)\langle n| = \sum_n |a_n\rangle\langle n|.$$

Similarly,

$$\mathbf{U}^\dagger = \mathbb{1}\mathbf{U}^\dagger = \sum_n |n\rangle\langle n|\mathbf{U}^\dagger = \sum_n |n\rangle\langle a_n|.$$

$$\mathbf{U}\mathbf{U}^\dagger = \sum_{nm} |a_n\rangle\langle n|m\rangle\langle a_m| = \sum_n |a_n\rangle\langle a_n| = \mathbb{1}.$$

(Same for $\mathbf{U}^\dagger\mathbf{U}$.) It is unitary.

Finally, I would like to explain the statement "We can diagonalize a Hermitian operator by a unitary transformation." According to the previous discussion, this is the same as saying that we can find a basis where a Hermitian operator is diagonal.

Suppose given a Hermitian operator $\mathbf{A} = \mathbf{A}^\dagger$. And suppose we are suffering along in some random basis $\{|n\rangle, n = 1..N\}$ in which $\mathbf{A}$ looks like

$$\mathbf{A} = \sum_{nm} |n\rangle\langle m| A_{nm}$$

where $\exists n \neq m$ such that $A_{nm} \neq 0$, *i.e.* A is not diagonal. Now consider the eigenvectors of $\mathbf{A}$, $\mathbf{A}|a_n\rangle = a_n|a_n\rangle$; we can choose $\{|a_n\rangle\}$ so that $\langle a_n|a_m\rangle = \delta_{nm}$ they form an orthonormal basis of $\mathcal{H}$.[8] What are the matrix elements of $\mathbf{A}$ in the $\{|a_n\rangle\}$ basis?

$$\langle a_r|\mathbf{A}|a_p\rangle = \delta_{rp} a_p$$

this is a diagonal matrix. And how are these matrix elements related to the ones in the other basis?

$$\mathbf{A} = \sum_{nm} \mathbb{1}|n\rangle\langle m|\mathbb{1} = \sum_{nmps} A_{nm}|a_p\rangle \underbrace{\langle a_p|n\rangle}_{\left(\mathbf{U}^\dagger\right)_{pn}} \underbrace{\langle m|a_r\rangle}_{\mathbf{U}_{mr}} \langle a_r| = \sum_{pr} \left(U^\dagger A U\right)_{pr} |a_p\rangle\langle a_r|$$

where I've left the matrix multiplication implicit in the last expression. We've shown that

$$\left(U^\dagger A U\right)_{pr} = \delta_{rp} a_p$$

is diagonal.

• The *rank* of a matrix or linear operator is the dimension of the space of states that it doesn't kill. By 'kill' I mean give zero when acting upon. The subspace spanned by vectors killed by $\mathbf{A}$ is called the *kernel* of $\mathbf{A}$. For an operator $\mathbf{A}$ acting on an N-dimensional Hilbert space (representable by an $N \times N$ matrix), $\text{rank}(A) = N-$ the dimension of the kernel of $\mathbf{A}$.

An invertible matrix has no kernel (you can't undo an operation that gives zero), and hence rank N. A matrix with rank less than N has zero determinant. (The determinant of $\mathbf{A}$ is the product of its eigenvalues: $\det \mathbf{A} = \prod_n a_n$, so vanishes if any of them vanish.)

A matrix like $|n\rangle\langle n|$ has rank 1: it only acts nontrivially on the one-dimensional space of states spanned by $|n\rangle$.

1.4 Axioms, continued

Important Conclusion. The important conclusion from the previous discussion is the following (I put it here because it is physics, not math). In quantum mechanics, the choices

[8] I say "can choose" because: (1) the normalization is not determined by the eigenvalue equation; (2) if there is a *degeneracy* in the spectrum ($\exists n \neq m$ such that $a_n = a_m$), then it is up to us to make an orthonormal basis for this subspace (e.g. by the Gram-Schmidt process)).

of basis come from observables. The labels that are supposed to go inside the little kets are possible values of observables – eigenvalues of Hermitian operators. This is why this feature of Dirac's notation – that we can put whatever we want in the box – is important: there are many possible observables we may consider diagonalizing in order to use their eigenbasis as labels on our kets.

Wavefunctions and choice of basis. An important question for you before we continue with the other three QM axioms: does this seem like the same notion of states that you are used to so far, where you describe the system by a *wavefunction* that looks something like $\psi(x)$? How are these the same quantum mechanics?! It took people (specifically, Dirac and Schrödinger) a little while to figure this out, and it will occupy us for the next few paragraphs. I've put this stuff in a different color, because it's not an essential part of the statement of the axioms of QM (though it is an essential part of relating them to the QM you know so far!).

To explain this, first recall the way you are used to thinking of a vector (*e.g.* in 3-space) as a list of numbers, indicating the *components* of the vector along a set of conventionally-chosen axes, which we could denote variously as:

$$\vec{x} = (x, y, z) = x\hat{x} + y\hat{y} + z\hat{z} = \sum_{i=1}^{3} x_i e_i = \sum_{i=1}^{3} x_i |i\rangle.$$

Here $\hat{x}, \hat{y}, \hat{z}$ or e_1, e_2, e_3 or $|1\rangle, |2\rangle, |3\rangle$ are meant to denote unit vectors point along the conventially-chosen coordinate axes. This set of unit vectors represents a **choice of basis** for the vector space in question (here $\mathbb{R}^3$); notice that this is extra data beyond just the definition of the vector space (and its inner product).

Now, consider the case of a free particle in one dimension. I claim that the data of its wavefunction $\psi(x)$ are the *components* of the associated vector in Hilbert space, in a particular basis. Specifically, in the basis where the position operator $\hat{x}$ is diagonal, whose basis vectors (like e_i) are labelled by a value of the position x, so an arbitrary vector in this vector space (which I will succumb to calling ψ) is a linear combination of vectors $|x\rangle$, with complex coefficients $\psi(x)$, like:

$$|\psi\rangle = \int dx\ \psi(x)|x\rangle\ .$$

We can project this equation onto a particular direction $\underline{x}$ by taking the inner product with $\langle \underline{x}|$, using $\langle \underline{x}|x\rangle = \delta(x - \underline{x})$:

$$\langle \underline{x}|\psi\rangle = \psi(\underline{x})\quad .$$

We can rewrite the same state in another basis, say the one where the momentum operator is diagonal, by inserting the relevant resolution of the identity:

$$\mathbb{1} = \int dp|p\rangle\langle p|$$

so that

$$|\psi\rangle = \int dx\ \psi(x)\mathbb{1}|x\rangle = \int dx\ \psi(x)\int dp|p\rangle\langle p|x\rangle \quad . \tag{2}$$

To make use of this, we need to know the overlaps of the basis states, which in this case are[9]:

$$\langle p|x\rangle = \frac{1}{\sqrt{2\pi\hbar}}e^{-ipx/\hbar} \ .$$

This says that the components in the momentum basis are the Fourier transform of those in the x-basis, since (2) is

$$|\psi\rangle = \int dp\left(\int \frac{dx}{\sqrt{2\pi\hbar}}e^{-ipx/\hbar}\psi(x)\right)|p\rangle \equiv \int dp\ \tilde{\psi}(p)|p\rangle.$$

So notice that this example of the QM of a particle, even in one dimension, is actually a very complex situation, since it hits you right away with the full force of Fourier analysis. Much simpler examples obtain if we instead think about finite-dimensional Hilbert spaces.

[End of Lecture 2]

Axiom 3: **Dynamics**

By 'dynamics' I mean dependence on time. Time evolution of a quantum state is implemented by the action of the *Hamiltonian* of the system, via

$$\frac{d}{dt}|\psi(t)\rangle = -\mathbf{i}\frac{\mathbf{H}}{\hbar}|\psi(t)\rangle \tag{3}$$

with $\mathbf{i} \equiv \sqrt{-1}$. To first order in a tiny time-interval, dt, this says:

$$\psi(t+dt) = (1 - \mathbf{i}\frac{\mathbf{H}}{\hbar}dt)\psi(t)$$

(this is the way you would write it, if you were going to implement this evolution numerically). [10]

---

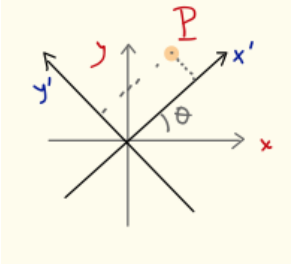
[9] You may wonder why this equation is true. In lecture, I was relying on your familiarity with the quantum description of the particle on a line, and didn't derive it. A quick way to derive this statement is to consider $\langle p|\hat{p}|x\rangle$ – one the one hand, $\hat{p} = \hat{p}^\dagger$ is acting to the left on its eigenstate, on the other hand (because of $[\hat{x},\hat{p}] = i\hbar$) it acts in the position basis as $-i\hbar\partial_x$. So the position-space components of $|p\rangle$ satisfy the following differential equation

$$-i\hbar\partial_x\langle x|p\rangle = p\langle x|p\rangle$$

whose solution is (up to a constant) the given expression. To fix the constant, consider

$$\delta(x_1 - x_2) = \langle x_1|x_2\rangle = \langle x_1|\mathbb{1}|x_2\rangle = \int dp\langle x_1|p\rangle\langle p|x_2\rangle.$$

[10] Notice that only the combination $\frac{\mathbf{H}}{\hbar}$ appears. Below I will just write $\mathbf{H}$. Similarly, instead of $\frac{\mathbf{p}}{\hbar}$ I will just write $\mathbf{p}$. Please see the scomment on units below.



Note that the preceding discussion of change of basis from position to momentum is structurally identical to a familiar basis rotation in ordinary space: we can label the coordinates of a point P in $\mathbb{R}^n$ ($n = 2$ in the figure) by its components along any set of axes we like. They will be related by: $x'_i = R^j_i x_j$ where in this case

$$R = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \quad \text{or} \quad R^j_i = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}^j_i = \langle j'|i\rangle$$

is the matrix of overlaps between elements of the primed and unprimed bases. So: using $\mathbb{1} = \sum_j |j'\rangle\langle j'|$, any vector P in $\mathbb{R}^n$ is

$$|P\rangle = \sum_i P^i |i\rangle = \sum_i P^i \left(\sum_j |j'\rangle\langle j'| \right) |i\rangle = \sum_j P^i R^j_i |j'\rangle \ .$$

The only difference between this simple rotation in 2-space and our previous discussion is that in Hilbert space we have to work over the complex numbers. In the case with real coefficients, the rotation R is a special case of a unitary matrix, called an *orthogonal* matrix with $R^t R = \mathbb{1}$ – we don't need to complex conjugate.

The operator

$$\mathbf{U}(dt) \equiv 1 - \mathbf{i}\mathbf{H}dt$$

which generates time evolution by the step dt, is unitary, because $\mathbf{H}$ is self-adjoint (up to terms that are small like dt^2):

$$\mathbf{U}^\dagger \mathbf{U} = \mathbb{1} + \mathcal{O}(dt^2) \ .$$

Unitary is important because it means that it preserves the lengths of vectors.

Successive application of this operator (if $\mathbf{H}$ is time-independent) gives

$$\mathbf{U}(t) = e^{-\mathbf{i}t\mathbf{H}} \ .$$

(Recall that $\lim_{N\to\infty}(1 + \frac{x}{N})^N = e^x$.) Notice that this solves the Schrödinger equation (3).

So: to evolve the state by a finite time interval, we act with a unitary operator:

$$|\psi(t)\rangle = \mathbf{U}(t)|\psi(0)\rangle \ .$$

Notice that energy eigenstates $\mathbf{H}|\omega\rangle = \hbar\omega|\omega\rangle$ play a special role in that their time evolution is very simple:

$$\mathbf{U}(t)|\omega\rangle = e^{-\mathbf{i}\frac{\mathbf{H}}{\hbar}t}|\omega\rangle = e^{-\mathbf{i}\omega t}|\omega\rangle$$

they evolve just by a phase. The evolution operator is diagonal in the same basis as $\mathbf{H}$ (indeed: $[\mathbf{H}, e^{i\mathbf{H}t}] = 0$):

$$\mathbf{U} = \sum_{\omega} e^{-\mathbf{i}\omega t}|\omega\rangle\langle\omega| \ .$$

Axiom 4: **Measurement**

This one has two parts:

(a) In quantum mechanics, the answer you get when you measure an observable $\mathbf{A}$ is an eigenvalue of $\mathbf{A}$.

(b) The measurement affects the state: right after the measurement, the system is in an eigenstate of $\mathbf{A}$ with the measured eigenvalue.

More quantitatively: if the quantum state just before the measurement were $|\psi\rangle$, then outcome a is obtained with probability

$$\mathbf{Prob}_\psi(a) = \| \mathbf{P}_a|\psi\rangle \|^2 = \langle\psi|\mathbf{P}_a|\psi\rangle \quad .$$

In the case where the eigenvalue a is non-degenerate, we have $\mathbf{P}_a = |a\rangle\langle a|$ and

$$\mathbf{Prob}_\psi(a) = \langle\psi|a\rangle\langle a|\psi\rangle = |\langle a|\psi\rangle|^2.$$

[End of Lecture 3]

If we get the outcome a, the quantum state becomes not $\mathbf{P}_a|\psi\rangle$ (which is not normalized!) but

$$|\psi\rangle \quad \overset{\text{measure } \mathbf{A}\text{, get } a}{\longrightarrow} \quad \frac{\mathbf{P}_a|\psi\rangle}{(\langle\psi|\mathbf{P}_a|\psi\rangle)^{1/2}} \tag{4}$$

which is normalized. (Check that if a is non-degenerate, (4) reduces to the simpler statement: $|\psi\rangle \overset{\text{measure } \mathbf{A}\text{, get } a}{\longrightarrow} |a\rangle$.) Notice that if we do the measurement again right away, the rule tells us that we are going to get the same answer, with probability one.

Some comments:

- Notice that spectral decomposition leads to the familiar expression for expectation values:

$$\langle\mathbf{A}\rangle \equiv \langle\psi|\mathbf{A}|\psi\rangle = \langle\psi|\sum_n a_n\mathbf{P}_n|\psi\rangle = \sum_n a_n\langle\psi|\mathbf{P}_n|\psi\rangle = \sum_n a_n\mathbf{Prob}_\psi(a_n) \ .$$

And notice that the fact that hermitian operators resolve the identity is crucial for the probabilistic interpretation: On the one hand

$$1 = \langle\mathbb{1}\rangle = \langle\psi|\psi\rangle = \| \, |\psi\rangle \|^2 \ .$$

On the other hand, for any $\mathbf{A} = \mathbf{A}^\dagger$, we can write this as

$$1 = \langle\psi|\psi\rangle = \langle\psi| \left(\sum_n \mathbf{P}_n\right) |\psi\rangle = \sum_n \mathbf{Prob}_\psi(a_n) \ .$$

Summing over all the probabilities has to give one.

- In light of the probabilistic interpretation in the measurement axiom, it makes sense that we want time evolution to happen via the action of a unitary operator, since the total probability (the probability that *something* will happen, including nothing as a possibility), had better always be 1, and this is equal to the magnitude of the state.

- Notice that while a sum of observables is an observable[11], a product of observables $\mathbf{AB}$ is not necessarily itself an observable, since[12]

$$(\mathbf{AB})^\dagger = \mathbf{B}^\dagger \mathbf{A}^\dagger \ . \tag{5}$$

  That is, $\mathbf{AB}$ is only self-adjoint if $\mathbf{A}$ and $\mathbf{B}$ commute, $[\mathbf{A}, \mathbf{B}] \equiv \mathbf{AB} - \mathbf{BA} = 0$. If the two operators may be simultaneously diagonalized, then by the measurement axiom, we can measure them simultaneously. So in that case, when the order of operation does not matter, it makes sense to think about the (unique) measurement of the product of the two.

- You may notice a glaring difference in character of our Measurement Axiom – all the others, in particular time evolution, involve *linear* operations:

$$\hat{A}(|a\rangle + |b\rangle) = \hat{A}|a\rangle + \hat{A}|b\rangle.$$

  Measurement, as described here (by the map labelled " $\xrightarrow{\text{measure } \mathbf{A}, \text{ get } a}$ " in (4) above) fails this property: it doesn't play nicely with superpositions. On the other hand, we think that the devices that we use to measure things (*e.g.* our eyeballs) are governed by quantum mechanics, and evolve in time via the (linear!) Schrödinger equation!

  We are going to have to revisit this issue. You might think that parts of Axiom 4 could be derived from a better understanding of how to implement measurements. You would not be alone.

- These axioms haven't been derived from something more basic, and maybe they can't be. In particular, the introduction of probability is an attempt to describe the results of experiments like particles going through a double slit, where interference patterns are observed. We don't know any way to predict what any one particle will do, but we can use this machinery to construct a probability distribution which describes with exquisite accuracy what many of them will do. And that machinery seems to apply to any experiment anyone has ever done.

- Comment on $\hbar$ and other unit-conversion factors: I am going to forget it sometimes. It is just a quantity that translates between our (arbitrary) units for frequency (that is, time^{-1}) and our (arbitrary) units for energy. We can choose to use units where it is one, and measure energy and frequency in the same units. Quantum mechanics is telling us that we should do this. If you need to figure out where the $\hbar$s are supposed to go, just think about which quantities are frequencies and which quantities are energies, and put enough $\hbar$s to make it work.

[11] $(\mathbf{A}+\mathbf{B})^\dagger = \mathbf{A}^\dagger + \mathbf{B}^\dagger = \mathbf{A} + \mathbf{B}$

[12] To see this: From the definition of adjoint, if $\mathbf{A}|u\rangle = |v\rangle$, then $\langle u|\mathbf{A}^\dagger = \langle v|$. So: $\mathbf{BA}|u\rangle = \mathbf{B}|v\rangle$, and so $\langle v|\mathbf{B}^\dagger = \langle u|\mathbf{A}^\dagger\mathbf{B}^\dagger$. But this is true for all $|u\rangle$, and we conclude (5).

- While I am warning you about possibly-unfamiliar conventions I use all the time without being aware of it: I may also sometimes use the Einstein summation convention without mentioning it. That is, repeated indices are summed, unless otherwise stated.

  Also, I will write $\partial_x \equiv \frac{d}{dx}$ because it involves fewer symbols and means the same thing.

1.5 Symmetries and conservation laws

[Preskill notes, section 2.2.1, Le Bellac section 8.1]

A fun and important thing we can derive already (to get used to the notation and because it is super-important) is the connection between (a) symmetries of a system and (b) quantities which don't change with time (conservation laws).

By *symmetry* I mean what you think I mean: some operation we can do to the system which doesn't change things we can observe about it. We've just finished saying that things we can observe in QM are eigenvalues of self-adjoint operators, and if we measure such an operator $\mathbf{A}$ in a state ψ, we get the outcome $|a\rangle$ with probability $|\langle a|\psi\rangle|^2$. So a symmetry is an operation on the system which preserves these probabilities. (We would also like it to respect the time evolution. More below.)

'Operation on the system' means a map on the states, that is on the vectors of the Hilbert space: $|\psi\rangle \to |\psi'\rangle$. And we would like to consider such operations which preserve the associated probabilities:

$$|\langle\phi|\psi\rangle|^2 = |\langle\phi'|\psi'\rangle|^2$$

for all ϕ, ψ. We can implement any such transformation by a linear and unitary operator[13] so that the symmetry acts by

$$|\psi\rangle \to |\psi'\rangle = \mathbf{U}|\psi\rangle$$

where $\mathbf{U}$ is unitary.

It is useful to notice that symmetries form a *group*, in the mathy sense: the product of two symmetries is a symmetry, and each one can be inverted. For each symmetry operation R on a physical system, there is a unitary operator $\mathbf{U}(R)$. The abstract group has a multiplication law, which has to be respected by its representation on the quantum states: first applying

[13] or *anti*-unitary operator. This innocuous-seeming statement that any symmetry can be represented this way is actually a theorem due to Wigner, which worries carefully about the fact that a map on the *rays* can be turned into a map on the *vectors*. Don't worry about this right now.

An "anti-unitary" transformation is one which is anti-linear and unitary, that is: $\phi_i \to \phi_i'$ with

$$\langle\phi_i'|\phi_j'\rangle = \langle\phi_j|\phi_i\rangle = \langle\phi_i|\phi_j\rangle^\star.$$

The anti-unitary case is only relevant for transformations which involve *time-reversal*. It is important for discrete symmetries but not for continuous ones. If you must, see Appendix A of Le Bellac for more information about the points in this footnote.

R_1 and then applying R_2 should have the same effect as applying their group product $R_2 \circ R_1$ (notice that they pile up to the left, in reverse lexicographic order, like in Hebrew). This means we have to have

$$\mathbf{U}(R_2)\mathbf{U}(R_1) \stackrel{?}{=} \mathbf{U}(R_2 \circ R_1) \ .$$

Actually, we can allow the slight generalization of this representation law:

$$\mathbf{U}(R_2)\mathbf{U}(R_1) = u(R_2, R_1)\mathbf{U}(R_2 \circ R_1) \ .$$

Here the object u is a (sometimes important) loophole: states are only defined up to an overall phase, so the group law only needs to be true up to such a phase $u(R_2, R_1)$ (this is called a *projective representation*).

Now the dynamics comes in: respecting the time evolution means that we should get the same answers for observables if we first do our symmetry operation and then evolve it in time (aka wait), or if we wait first and then do the operation. That is we have to have:

$$\mathbf{U}(R)e^{-\mathbf{i}t\mathbf{H}} = e^{-it\mathbf{H}}\mathbf{U}(R). \tag{6}$$

Expanding out (6) to linear order in t we have

$$\mathbf{U}(R)\mathbf{H} = \mathbf{H}\mathbf{U}(R) \qquad i.e. \quad [\mathbf{U}(R), \mathbf{H}] = 0. \tag{7}$$

For a continuous symmetry (like rotations or translations) we can say more. We can choose R to be very close to doing nothing, in which case $\mathbf{U}$ must be close to the identity on the Hilbert space

$$\mathbf{U} = \mathbb{1} - \mathbf{i}\epsilon\mathbf{Q} + \mathcal{O}(\epsilon^2) \ .$$

$\mathbf{U}$ is unitary; this implies that $\mathbf{Q}^\dagger = \mathbf{Q}$, $\mathbf{Q}$ is an observable. Expanding (7) to first order in ϵ, we find

$$[\mathbf{Q}, \mathbf{H}] = 0 \ . \tag{8}$$

This is a *conservation law* in the following sense: if the system is in an eigenstate of $\mathbf{Q}$, the time evolution by the Schrödinger equation doesn't change that situation. Symmetries imply conservation laws.

Conversely, given a conserved quantity $\mathbf{Q}$ (*i.e.* an observable satisfying (8)), we can construct the corresponding symmetry operations by

$$\mathbf{U}(R) = \lim_{N\to\infty}\left(\mathbb{1} - \mathbf{i}\frac{\theta}{N}\mathbf{Q}\right)^N = e^{-i\theta\mathbf{Q}}.$$

The conserved quantity $\mathbf{Q}$ is the *generator* of the symmetry, in the sense that it generates an infinitesimal symmetry transformation.

As an example, let's think about a free particle that lives on a line. Its Hilbert space is labelled by a position x, with $x \in [-\infty, \infty]$. When I say 'free' particle, I mean that the Hamiltonian is

$$\mathbf{H} = \frac{\hat{p}^2}{2m} = -\frac{\hbar^2}{2m}\partial_x^2 \ .$$

Notice that shifting the origin of the line doesn't change the Hamiltonian; formally, this is the statement that

$$[\hat{p}, \mathbf{H}] = 0 \ .$$

The fact that the Hamiltonian doesn't depend on x means that momentum is a conserved charge. What is the finite transformation generated by $\hat{p}$? It is just $\mathbf{U}(a) = e^{-ia\hat{p}}$ which acts on a position-basis wavefunction $\psi(x) = \langle x|\psi\rangle$ by

$$\mathbf{U}(a)\psi(x) = e^{-ia\hat{p}}\psi(x) = e^{a\hbar\partial_x}\psi(x) = \psi(x) + a\hbar\partial_x\psi(x) + \frac{1}{2!}(a\hbar)^2\partial_x^2\psi(x) + ... = \psi(x + a\hbar) \ ,$$

a translation. Here I used Taylor's theorem.

The group law in this example is very simple: $\mathbf{U}(a)\mathbf{U}(b) = \mathbf{U}(a+b) = e^{i(a+b)\hat{p}} = \mathbf{U}(b)\mathbf{U}(a)$. This group is *abelian*: all the elements commute.

Below we will see another example where the symmetry group is non-abelian, and where the projective loophole is exploited.

[End of Lecture 4]

One final comment about symmetry operations in general: we've shown that a symmetry R is implemented on $\mathcal{H}$ by a unitary operator $\mathbf{U}(R)$ acting on the states:

$$|\psi\rangle \to \mathbf{U}(R)|\psi\rangle \ .$$

Recall that unitaries implement transformations between ON bases. This tells us how they should act on operators:

$$\mathbf{A} \to \mathbf{U}(R)\mathbf{A}\mathbf{U}(R)^\dagger$$

This guarantees that expressions like expectation values are unchanged:

$$\langle\phi|\mathbf{A}|\psi\rangle \to \langle\phi|\mathbf{U}^\dagger\left(\mathbf{U}\mathbf{A}\mathbf{U}^\dagger\right)\mathbf{U}|\psi\rangle = \langle\phi|\mathbf{A}|\psi\rangle.$$

Notice that our statement that the time evolution operator commutes with the symmetry operation (6) can be rewritten as:

$$\mathbf{U}(R)e^{-\mathbf{i}t\mathbf{H}}\mathbf{U}(R)^\dagger = e^{-\mathbf{i}t\mathbf{H}}$$

the action of the symmetry preserves the time-evolution operator (and the Hamiltonian).

In conclusion here, the real practical reason we care about symmetries is that they allow us to solve problems (just like in classical mechanics). In particular, the fact that the symmetry operator $\mathbf{Q}$ commutes with $\mathbf{H}$ means that $\mathbf{Q}$ eigenstates are energy eigenstates. If we can diagonalize $\mathbf{Q}$, we can (more likely) find the spectrum of the Hamiltonian, which is what we mean by solving a QM system.

1.6 Two-state systems: Pauli matrix boot camp

[Preskill notes, section 2.2; Le Bellac, chapter 3]

The smallest non-trivial Hilbert space is two-dimensional. If we call the generators of an orthonormal basis $|\uparrow\rangle$ and $|\downarrow\rangle$ then any normalized state can be written as

$$z|\uparrow\rangle + w|\downarrow\rangle \tag{9}$$

with $|z|^2 + |w|^2 = 1$. The set of complex numbers $\{(z, w) \text{ s.t. } |z|^2 + |w|^2 = 1\}$ describes a three-sphere S^3; but we must remember that the overall phase of the state is not meaningful. The resulting space $\{(z, w) \in \mathbb{C}^2 \text{ s.t. } |z|^2 + |w|^2 = 1\}/\left((z, w) \simeq e^{i\varphi}(z, w)\right)$ is the projective plane $\mathbb{CP}^1$, aka a two-sphere[14]. In this context, where it parametrizes the states of a qbit, it is called the Bloch sphere.

You have seen this Hilbert space realized in the discussion of spin-1/2 particles, as a kind of inherent 'two-valuedness' of the electron (to quote Pauli). In that case, the two real coordinates on the Bloch sphere (conventionally, the polar angle θ and the azimuthal angle φ) represent the orientation of the spin of the particle, as we'll reconstruct below.

This same structure also represents any kind of quantum system with a two-dimensional Hilbert space, for example: polarizations of a photon, energy levels of a two-level (approximation to an) atom, the two configurations of an ammonia molecule, two possible locations of an electron in a H_2^+ ion ... The perhaps-overly-fancy quantum information theory name for such a system is a 'qbit' (short for 'quantum bit,' sometimes spelled 'qubit'.).

Our measurement axiom tells us the interpretation of $|z|^2$ and $|w|^2$ in the state (9). Next I want to talk about the physical significance of the relative phase.

1.6.1 Symmetries of a qbit

[Le Bellac 8.2.3] Let's talk about rotations (and therefore angular momentum), briefly. Partly I am doing this as the promised example of a non-abelian symmetry, and partly because it will be useful for further thinking about qbits. We are used to thinking about rotations of 3-space acting on a vector, $v_i' = \sum_j R_i^j v_j$. To specify which rotation we are talking about we need to specify an axis of rotation $\hat{n}$ and an angle of rotation θ. An infinitesimal rotation by $d\theta$ about the axis $\hat{n}$ looks like

$$R(\hat{n}, d\theta) = \mathbb{1} - \mathbf{i} d\theta \hat{n} \cdot \vec{J} \tag{10}$$

[14] It is more clear that this is an S^2 if we think of it as

$$\{(z, w) \in \mathbb{C}^2\}/\left((z, w) \simeq \lambda(z, w), \lambda \in \mathbb{C} \setminus \{0\}\right);$$

a good label on equivalence classes is z/w which is an arbitrary complex number. The only catch is that $w = 0$ is a perfectly good point; it is the point at infinity in the complex plane.

where $\vec{J}$ is the angular momentum vector, the *generator* of rotations (recall the general discussion just above: the conserved charge is the generator of the symmetry).

Acting on vectors, $\vec{J}$ is three three-by-three matrices; you can convince yourself that you'll get the answers you expect by taking

$$(J^i)^k_j = \mathbf{i}\epsilon_{ijk}$$

where ϵ_{ijk} is the completely antisymmetric object with $\epsilon^{123} = 1$ [15]. For example,

$$J^z = \mathbf{i}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} .$$

So the infinitesimal rotation (10) with $\hat{n} = \hat{z}$ turns $\hat{x}$ into $\hat{x}$ plus a little bit of $\hat{y}$. Notice that the factors of $\mathbf{i} \equiv \sqrt{-1}$ cancel out here.

A finite rotation is

$$R(\hat{n}, \theta) = e^{-\mathbf{i}\theta\hat{n}\cdot\vec{J}} .$$

Rotations about distinct axes don't commute. The algebra of generators (which you can discover by thinking about doing various infinitesimal rotations in different orders) is[16]:

$$[J_j, J_k] = \mathbf{i}\epsilon_{jkl}J_l . \tag{11}$$

So far, we've been discussing the action of rotations on a vector or spin-1 object. Next we think about spin-$\frac{1}{2}$. The Pauli spin matrices are defined by universally-agreed convention to be:

$$\boldsymbol{\sigma}^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \boldsymbol{\sigma}^y = \begin{pmatrix} 0 & -\mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix} \quad \boldsymbol{\sigma}^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

(occasionally I will write $\boldsymbol{\sigma}^x \equiv \boldsymbol{\sigma}^1, \boldsymbol{\sigma}^y \equiv \boldsymbol{\sigma}^2, \boldsymbol{\sigma}^z \equiv \boldsymbol{\sigma}^3$). You will demonstrate various of their virtues on the second homework, including the fact that they are Hermitian (and hence observables of a qbit). You will also show that

$$\boldsymbol{\sigma}^i\boldsymbol{\sigma}^j = \mathbf{i}\epsilon^{ijk}\boldsymbol{\sigma}^k + \delta^{ij}\mathbb{1} \tag{12}$$

and in particular that the objects

$$\mathbf{J}^{\left(\frac{1}{2}\right)}_k \equiv \frac{1}{2}\boldsymbol{\sigma}_k$$

satisfy the angular momentum algebra (11). I've put the $\left(\frac{1}{2}\right)$ superscript to emphasize that this is a particular representation of the algebra (11), the spin-1/2 representation. Notice

[15] that is: $\epsilon_{ijk} = 0$ if any of ijk are the same, $= 1$ if ijk is a cyclic permutation of 123 and $= -1$ if ijk is an odd permutation of 123, like 132.

[16] Cultural remark: this algebra is called $\mathbf{so}(3)$ or $\mathbf{su}(2)$.

that the factor of 1/2 in the relation between $\mathbf{J}$ and $\boldsymbol{\sigma}$ is significant for getting the right factor on the RHS of (11). The fact that the $\boldsymbol{\sigma}$s are Hermitian means that this object is an observable acting on the Hilbert space of our qbit.

Now we follow our procedure for getting finite transformations from symmetry generators: The matrix which implements a finite rotation of an object whose angular momentum is $\vec{\mathbf{J}}^{(\frac{1}{2})} = \frac{1}{2}\vec{\boldsymbol{\sigma}}$ by an angle θ about the axis $\hat{n}$ is then:

$$\mathbf{U}(R_{\hat{n}}(\theta)) = e^{-\mathbf{i}\theta\hat{n}\cdot\vec{J}^{1/2}} = e^{-\mathbf{i}\frac{\theta}{2}\vec{\boldsymbol{\sigma}}\cdot\hat{n}} \ .$$

In the homework you'll show that this is:

$$e^{-\mathbf{i}\frac{\theta}{2}\vec{\boldsymbol{\sigma}}\cdot\hat{n}} = \mathbb{1}\cos\frac{\theta}{2} - \mathbf{i}\vec{\boldsymbol{\sigma}}\cdot\hat{n}\sin\frac{\theta}{2} \tag{13}$$

where $\hat{n}$ is a unit vector, using the Taylor expansion (really the definition) of the exponential.

Notice that $\mathbf{U}(R(2\pi)) = -\mathbb{1}$. (!) A rotation by 2π does nothing to a vector. But it multiplies a spinor (the wave function of a spin-half object) by a phase. This is a projective representation of the rotation group SO(3). [17]

Don't be tricked by the fact that this appears in the phase of the state into thinking that this minus sign isn't important. When there is more than one qbit in the world (below!), it will matter.

Notice the remarkable fact that we can have a *continuous* symmetry acting on a system with just two states. This cannot happen in classical mechanics!

[End of Lecture 5]

Other components of the spin

The point of what we are about to do is to give a physical interpretation of the relative phase in the superposition of $\boldsymbol{\sigma}^z$ eigenstates (9). There are some big expressions, so I'll give away the ending: the punchline is that it encodes the spin along the *other* axes, $\boldsymbol{\sigma}^{x,y}$.

More generally, it is often useful to have an explicit expression for the eigenvectors of $\hat{n}\cdot\vec{\boldsymbol{\sigma}}$ in the $\boldsymbol{\sigma}^z$ eigenbasis. Here is a slick way to get them (easier than brute force).

Let R be a rotation (about some axis $\hat{n}$, by some angle θ, which I won't write). If I rotate my reference axes, R is still a rotation by an amount θ, but about a differently-labelled axis. This innocuous-seeming observation implies the following beautiful equation:

$$\mathbf{U}(R)\mathbf{J}_k\mathbf{U}(R)^\dagger = R^l_k\mathbf{J}_l = (R\mathbf{J})_k$$

[17] It is actually an ordinary representation of the group SU(2) which is a double-cover of SO(3). Fancy math comment: the existence of such a projective representation is related to the fact that SO(3) is not simply connected.

which in words is the unsurprising statement that:

angular momentum is a vector

i.e. it transforms under rotations like a vector.[18]

So in particular, if $|m\rangle$ is an eigenstate of $\mathbf{J}^z$

$$\mathbf{J}_z|m\rangle = m|m\rangle$$

then its image under the rotation $\mathbf{U}(R)|m\rangle$ is an eigenstate of the image of $\mathbf{J}^z$, $R\mathbf{J}^z$, with the same eigenvalue:

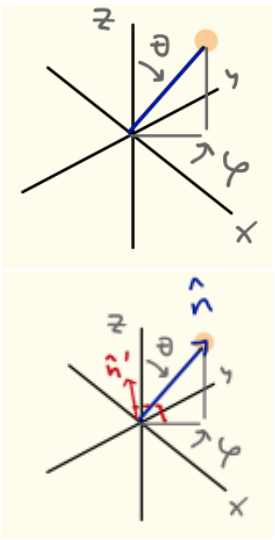
$$(R\mathbf{J})_z\left(\mathbf{U}(R)|m\rangle\right) = \mathbf{U}(R)\mathbf{J}_z\mathbf{U}(R)^\dagger\mathbf{U}(R)|m\rangle = \mathbf{U}(R)\mathbf{J}_z|m\rangle = m\left(\mathbf{U}(R)|m\rangle\right) \ .$$

So: we can construct eigenstates of the spin operator along other axes $\boldsymbol{\sigma}^n \equiv \hat{n}\cdot\vec{\boldsymbol{\sigma}}$ by applying an appropriate rotation to a $\mathbf{J}_z$ eigenstate. Which rotation?
[This is slippery! Think about it in a quiet place.]
Claim: we can take the vector $\hat{z}$ to the vector $\hat{n} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ by applying a rotation through angle θ about the axis $\hat{n}' = (-\sin\varphi, \cos\varphi, 0)$. Acting on our qbit, this is

$$e^{-\mathbf{i}\frac{\theta}{2}\hat{n}'\cdot\vec{\boldsymbol{\sigma}}} = \begin{pmatrix} \cos\frac{\theta}{2} & -e^{-\mathbf{i}\varphi}\sin\frac{\theta}{2} \\ e^{\mathbf{i}\varphi}\sin\frac{\theta}{2} & \cos\frac{\theta}{2}\end{pmatrix}$$

using (13).

The convention is: θ is the polar angle measured from the $+\hat{z}$ axis, and φ is the azimuthal angle from the $\hat{x}$ axis, like this:



That is, the eigenstates $\hat{n}\cdot\vec{\boldsymbol{\sigma}}|\pm,\hat{n}\rangle = \pm|\pm,\hat{n}\rangle$ are:

$$|+,\hat{n}\rangle = \begin{pmatrix} e^{-\mathbf{i}\varphi/2}\cos\frac{\theta}{2} \\ e^{+\mathbf{i}\varphi/2}\sin\frac{\theta}{2}\end{pmatrix} \ , \quad |-,\hat{n}\rangle = \begin{pmatrix} -e^{-\mathbf{i}\varphi/2}\sin\frac{\theta}{2} \\ e^{+\mathbf{i}\varphi/2}\cos\frac{\theta}{2}\end{pmatrix} \ . \tag{14}$$

The big conclusion of this discussion is that the vector in (9) with $z = e^{-i\varphi/2}\cos\frac{\theta}{2}, w = e^{+i\varphi/2}\sin\frac{\theta}{2}$, is a spin pointing in the $\hat{n}$ direction! This is the significance of the relative phase. For example, take $\theta = \pi/2$ and $\varphi = 0$ to get the states

$$\frac{1}{\sqrt{2}}\left(|\uparrow\rangle + |\downarrow\rangle\right) \equiv |\uparrow_x\rangle, \quad \frac{1}{\sqrt{2}}\left(|\uparrow\rangle - |\downarrow\rangle\right) \equiv |\downarrow_x\rangle$$

[18] You may want to tell me that angular momentum is a 'pseudo-vector'; this name refers to the fact that it is odd under a parity transformation and doesn't contradict my statement about how it behaves under a rotation.

As the names suggest, these are eigenstates of $\boldsymbol{\sigma}^x$; they represent spin up or down along the x direction. More generally we have shown that

$$|\uparrow_{\hat{n}}\rangle = e^{-\mathbf{i}\varphi/2}\cos\frac{\theta}{2}|\uparrow\rangle + e^{+\mathbf{i}\varphi/2}\sin\frac{\theta}{2}|\downarrow\rangle$$

(if I don't put a subscript on the arrows, it means spin along $\hat{z}$) is a spin pointing (up) in the direction $\hat{n}$.

1.6.2 Photon polarization as a qbit

[Preskill 2.2.2] Another important example of a physical realization of a qbit is the polarization states of a photon. I mention this partly because it is important in experimental implementations of the physics discussed below, and also to emphasize that despite the fact that any two-state system has the structure of an electron spin, they don't all have to transform the same way under rotations.

Recall that an electromagnetic plane wave (with fixed wavevector) has two independent polarization states, with $\vec{E}$ and $\vec{B}$ transverse to the wavevector. The same is true of the quantum mechanical version of this excitation, a photon. So a photon is a qbit that moves at the speed of light.

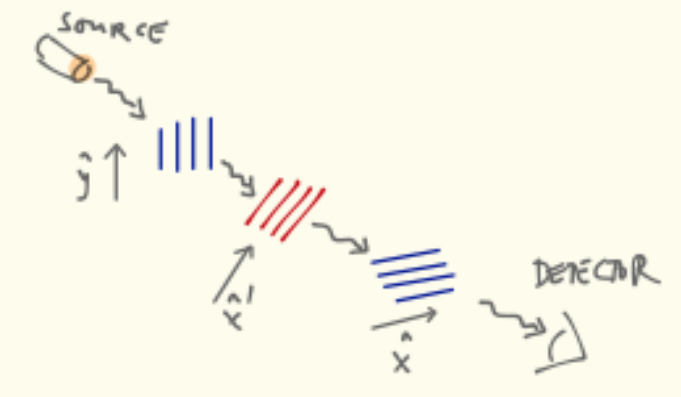
I should comment about how this qbit behaves differently than a spin-$\frac{1}{2}$ particle under rotations. These two states of a photon have to transform into each other only under rotations that preserve the wavevector of the photon; if the photon is headed in the $\hat{z}$ direction, this is just rotations in the xy-plane[19]. The two linear polarization basis states $|x\rangle$ and $|y\rangle$ rotate via the matrix: $\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}$, the ordinary rotation matrix acting on a two-dimensional vector – the photon's polarization is not a spinor.[20] **Interference** in this context is something that you've probably seen with your eyes (though you probably haven't seen it happen for single photons): you can measure various components of the polarization using polarizing filters (this is the analog of the Stern-Gerlach apparatus for electron spin). If we put crossed linear polarizers, no photons get through. But if we stick a linearly-polarizing filter tilted at 45° in between the crossed polarizers some photons (a quarter of them) get through again.

[19] Cultural remark: the fancy name for the subgroup of rotations that preserve the wavevector is the "little group"; this concept is useful for classifying representations of the Poincaré group.

[20] The generator of this rotation is $\boldsymbol{\sigma}^y$, with eigenvalues ± 1 (not $\pm\frac{1}{2}$), hence we can call the photon spin 1. The eigenstates of angular momentum are

$$|R\rangle \equiv \frac{1}{\sqrt{2}}\left(|x\rangle + \mathbf{i}|y\rangle\right), \quad |L\rangle \equiv \frac{1}{\sqrt{2}}\left(\mathbf{i}|x\rangle + |y\rangle\right)$$

which are called right- and left-handed circular polarization states. Beware that not everyone agrees which is which; it depends on whether you think the light is coming at you or going away from you.



To see this: Say the first filter is aligned along y – it projects the initial state onto $|y\rangle$. It is a *measurement* of the y-polarization of the photon. The final filter projects onto $|x\rangle$ which is orthogonal to $|y\rangle$, so if nothing happens in between, nothing gets through the crossed polarizers. But the filter tilted at 45° to the $\hat{x}$ axis projects the photon state onto its component along $|x'\rangle = \frac{1}{\sqrt{2}}(|x\rangle + |y\rangle)$. (Note that this is the analog of $|\uparrow_x\rangle = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle + |\downarrow_z\rangle)$.) Since we can decompose $|y\rangle = \frac{1}{\sqrt{2}}(|x'\rangle + |y'\rangle)$, and $|x\rangle = \frac{1}{\sqrt{2}}(|x'\rangle - |y'\rangle)$, we see that indeed something will get through.

Note that here we see the normalization step of the measurement axiom very vividly: the polarizers absorb some of the photons – they decrease the intensity of the light that gets through. But *if* we see a photon at the detector, the polarizer acts as a measurement of its polarization. This conditional probability – conditioned on the photon getting through – is the thing that needs to be normalized in this case. This is the sense in which the polarizer implements the measurement.

A brief word about quantum information

Suppose we have in our hands a qbit (say an electron) in the state (9) and let's think about trying to do measurements to determine the direction $\hat{n}$.

If we measure the spin along $\hat{z}$, according to Axiom 4, we get outcome $|\uparrow\rangle$ with probability $|z|^2$ and $|\downarrow\rangle$ with probability $|w|^2$, and after the measurement the state is *certainly* whichever answer we got. With a single copy of the state, we don't learn anything at all!

With many identical qbits prepared in this state, we could measure the probability distribution, by counting how many times we get each outcome. This would give us a measurement of $|z|^2 = 1 - |w|^2 = \cos^2\frac{\theta}{2}$. For this distribution, parametrized only by one variable θ, this is information we get by measuring the expectation value of the spin along $\hat{z}$:

$$\langle \boldsymbol{\sigma}^z \rangle = \langle \uparrow_{\hat{n}} | \boldsymbol{\sigma}^z | \uparrow_{\hat{n}} \rangle = \cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2} = \cos\theta.$$

Neither of these things tell us anything at all about φ!

To learn about φ, we would have to measure the spin along a different axis. But that would destroy the information about θ!

This is an important sense in which a qbit is really different from a probability distribution for a classical bit: because of the possibility of quantum interference, the probabilities don't add the same way. We'll have more to say about this.

[End of Lecture 6]

1.6.3 Solution of a general two-state system

[Feynman III-7-5, Le Bellac Chapter 4] Part of the reason for the ubiquity of the Pauli sigma matrices is the following two facts about operators acting on a two-state Hilbert space, which say that they can basically all be written as sums of Pauli matrices.

1. Any Hermitian 2-by-2 matrix $\mathbf{h}$ (such as any Hamiltonian of any two-state system) can be expanded as

$$\mathbf{h} = d_0 \mathbb{1} + \vec{\boldsymbol{\sigma}} \cdot \vec{d} = \begin{pmatrix} d_0 + d_3 & d_1 - \mathbf{i} d_2 \\ d_1 + \mathbf{i} d_2 & d_0 - d_3 \end{pmatrix} . \tag{15}$$

We can see this by counting: any 2×2 Hermitian matrix is of the form $\begin{pmatrix} a & b \\ b^\star & c \end{pmatrix}$ with a, c real; this is exactly (15) with $a = d_0 + d_3, c = d_0 - d_3, b = d_1 - \mathbf{i} d_2$.

We have already basically figured out the eigensystem for this matrix. There are two differences between (15) and $\vec{\boldsymbol{\sigma}} \cdot \hat{n}$: First, we add a multiple of the identity. But that doesn't change the eigenvectors at all: the identity commutes with everybody, and so can be simultaneously diagonalized with anybody. All this does is add d_0 to each eigenvalue. Second, $\vec{d}$ is not necessarily a unit vector; rather $\vec{d} = |d| \hat{n}$, where $|d| \equiv \sqrt{\vec{d} \cdot \vec{d}}$. But $\vec{d} \cdot \vec{\boldsymbol{\sigma}}$ and $\hat{n} \cdot \vec{\boldsymbol{\sigma}}$ also have the same eigenvectors, we are just multiplying the matrix by a number $|d|$. Combining these facts, the eigenvectors of *any* 2-by-2 Hermitian matrix $\mathbf{h}$ are of the form (14) with eigenvalues

$$\epsilon_\pm = d_0 \pm \sqrt{|d|}.$$

2. And any unitary matrix acting on a qbit is of the form

$$\mathbf{U} = e^{\mathbf{i}\varphi} e^{-\mathbf{i}\frac{\theta}{2}\hat{n}' \cdot \vec{\boldsymbol{\sigma}}}$$

for some θ and some $\hat{n}$, and where $e^{\mathbf{i}\varphi}$ is a phase. [21] Again we can see this by comparing to the most general form consistent with $\mathbf{U}\mathbf{U}^\dagger = \mathbb{1}$. This implies that any unitary operation

[21] Notice that the rotation matrix $e^{-\mathbf{i}\frac{\theta}{2}\hat{n}' \cdot \vec{\boldsymbol{\sigma}}}$ has unit determinant. One way to see this is to use the fact that for any matrix $\mathbf{M}$ (with eigenvalues $\{\lambda\}$),

$$\log \det \mathbf{M} = \log \prod_\lambda \lambda = \sum_\lambda \log \lambda = \operatorname{tr} \log \mathbf{M} .$$

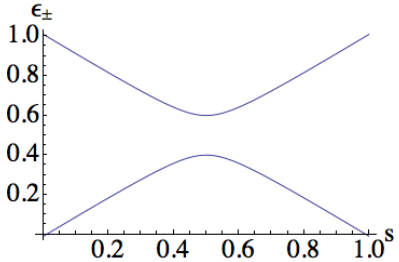
that you might want to do to a two-level system (*e.g.* the time evolution of a two-level system, as in HW 2) can be thought of (up to an often-irrelevant overall phase) as a rotation of the spin of the qbit.

Avoided level crossings

A simple application of the previous discussion demonstrates a phenomenon called *avoided level crossings*. Suppose we have a knob on our quantum system: the Hamiltonian $\mathbf{h}$ depends on a parameter I'll call s. There's a Hamiltonian for each s, each with its own eigenvalues and eigenvectors. I can draw a plot of the eigenvalues of $\mathbf{h}(s)$ as a function of s.

(You'll believe me that if I start the system in some eigenstate i of $\mathbf{h}(0)$ and I turn this knob slowly enough, then the system will follow the curve $\epsilon_i(s)$ – this is called the *adiabatic approximation*.)

Question: do you think these curves cross each other as s varies?



The answer is no, as we can see by the previous analysis. Without loss of generality (WLOG), we can focus on a pair of neighboring levels and ask if they will cross: this reduces $\mathbf{h}$ to acting on a qbit, so $\mathbf{h} = d_0 \mathbb{1} + \vec{d} \cdot \vec{\boldsymbol{\sigma}}$. But in order for the two levels of the qbit to collide, $\epsilon_+ = \epsilon_-$, we need to have $|d| = 0$, which means $\vec{d} = 0$, which is *three* equations. By varying one parameter, something special would have to happen in order run into a point where three equations were satisfied. So: *generically*, levels don't cross.

$$\mathbf{h}(s) = \begin{pmatrix} s & \epsilon \\ \epsilon & 1-s \end{pmatrix}$$

($\epsilon = .1$ in the plot).

Also, note that the crossing is avoided despite the fact that (in the example in the figure) when $s = 0$, the lower level is $\begin{pmatrix} 1 & 0 \end{pmatrix}$ and when $s = 1$, the lower level is $\begin{pmatrix} 0 & 1 \end{pmatrix}$ – the two states really did switch places!

(If there is some extra symmetry forcing some (two) of the components of $\vec{d}$ to vanish in our family of hamiltonians, we can force a crossing to occur.)

Finally, note that this result is general, not just for two-level systems. For an arbitrary-dimensional $\mathcal{H}$, we may focus on two adjacent energy levels which are in danger of crossing, and we are left with $\mathbf{h}$ again – including the couplings to more levels only introduces *more* parameters.

For the rotation matrix, this gives:

$$\det e^{-\mathbf{i}\frac{\theta}{2}\hat{n}'\cdot\vec{\boldsymbol{\sigma}}} = e^{\operatorname{tr}\log e^{-\mathbf{i}\frac{\theta}{2}\hat{n}'\cdot\vec{\boldsymbol{\sigma}}}} = e^{\operatorname{tr}\left(-\mathbf{i}\frac{\theta}{2}\hat{n}'\cdot\vec{\boldsymbol{\sigma}}\right)} = e^0 = 1.$$

since the $\boldsymbol{\sigma}$s are traceless.

1.6.4 Interferometers: photon trajectory as a qbit

[Schumacher §2.1] Let's think about another possible realization of simple (few-state) quantum systems using photons. In free space, light can go all over the place; it's easier to think about an *interferometer*, an apparatus wherein light is restricted to discrete *beams*, which can be guided or split apart and recombined. A beam is by definition a possible path for a photon.

We are going to think about *linear* optical devices which do not themselves create or destroy photons, and which are not disturbed by the passage of the photons.

Now: suppose we input a beam into an interferometer with two paths, which we'll call upper and lower, and a detector at the end. Then each beam will have an associated probability amplitude α, β and the probability that we would find the photon on the top path (if we were to put a detector there) is $P_{\text{upper}} = |\alpha|^2$. By superposition, can represent the intermediate state of the photon by

$$|\psi\rangle = \alpha|\text{upper}\rangle + \beta|\text{lower}\rangle \ .$$

If we know that we sent in a photon, then we must have

$$1 = P_{\text{upper}} + P_{\text{lower}} = |\alpha|^2 + |\beta|^2.$$

Notice that we are again encoding a qbit in a photon, but not in its polarization, but rather in its choice of path.

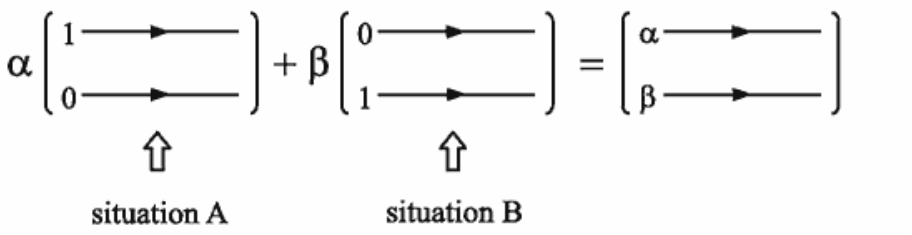


Figure 1: [From Schumacher] Here is an illustration of superposition.

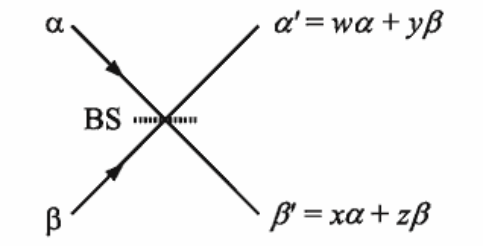
There are a few things we can do to our photons while we're sending them through our maze. One device we can send them through is a **phase shifter**. This could be a glass plate through which the light goes. In doing so, it gets slowed down, so its state acquires a larger phase $\sim i\omega t$. We know the photon went in and will come out, so the probability $|\alpha|^2$ is preserved, but its phase changes:

$$\alpha \mapsto e^{i\delta}\alpha$$

where δ is a property of the device. For example, it could be $\delta = \pi$, in which case this is is $\alpha \mapsto -\alpha$. This phase is important; for example, it can turn constructive interference into destructive interference.

Next, we can send our beam into a **beamsplitter**; this takes the input beam and splits it into two beams of lower intensity. A partially-silvered mirror will accomplish this by partially reflecting and partially transmitting the wave. Such a mirror has two possible basis inputs (see fig): light from above or below, which we can denote respectively by the vectors $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$. By the principle of superposition (the assumption that the device is linear), we must have that the action of the beamsplitter is

$$\begin{pmatrix}\alpha\\\beta\end{pmatrix} \mapsto \begin{pmatrix}\alpha'\\\beta'\end{pmatrix} = \mathbf{U}\begin{pmatrix}\alpha\\\beta\end{pmatrix} = \begin{pmatrix}w & y\\x & z\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix}.$$



If we know that all of the photons end up in one or the other of the two output beams, then the matrix $\mathbf{U}$ must be unitary – it must conserve probability.

For a half-silvered mirror, by definition the reflection and transmission probabilities are equal, and therefore must both be 1/2. Hence, we must have $|w| = |y| = |x| = |z| = \frac{1}{\sqrt{2}}$.
Question: Can we make the simplest-seeming choice of signs

$$\mathbf{U} \stackrel{?}{=} \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix} \quad ?$$

Well, it acts fine on the basis states; but what happens if we send in the perfectly good state $\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$? Then the output is

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \stackrel{?}{\mapsto} \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}$$

– the photon has probability $|1|^2 = 1$ for being found in each beam! The total probability grew. That matrix wasn't unitary.

A choice which does work is

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

which is called a 'balanced beamsplitter'. The minus sign means that when the lower input beam is reflected, it experiences a phase shift by π; this is necessary for conserving probability and would be realized if we had a good quantum mechanical model of the mirror. (It is also consistent with classical optics.) Notice that this means that the beamsplitter can't be symmetrical between up and down; the side with the phase shift is sometimes denoted with a dot.

In the same notation, a phase shifter which acts only on the upper leg of the interferometer acts by the matrix

$$\mathbf{P}_{\text{upper}}(\delta) = \begin{pmatrix}e^{i\delta} & 0\\0 & 1\end{pmatrix}.$$

By another device, we can let the beams cross, so that we exchange the upper and lower amplitudes; this is represented by

$$\mathbf{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \boldsymbol{\sigma}^x.$$

Finally, a fully-silvered mirror acts by a π-phase shift on the beam.

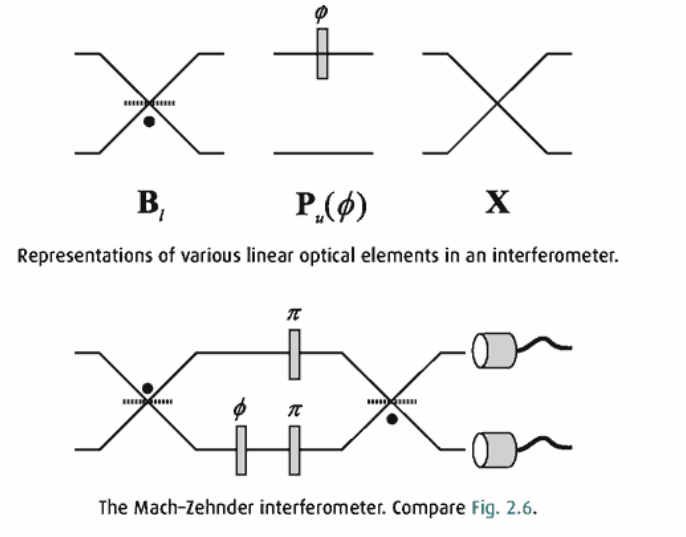


Figure 2: Interferometer ingredients, from Schumacher.

We are going to use these ingredients to do something with quantum mechanics that cannot be done classically in section 1.9.2.

1.7 Composite quantum systems

[Preskill section 2.3; Le Bellac chapter 6; Weinberg chapter 12]

1.7.1 Tensor products (putting things on top of other things)

[Schumacher section 6.1, Le Bellac section 6.1] Here we have to introduce the notion of combining quantum systems. Suppose we have two spin 1/2 particles. We will fix their locations, so that their position degree of freedom does not fluctuate, and so that they are distinguishable (later we will see that quantum statistics complicates the rule I am about to state). Each is described by a two-dimensional Hilbert space, a qbit:

$$\mathcal{H}_{a=1,2} = \text{span}\{|\uparrow\rangle_a, |\downarrow\rangle_a\} \ .$$

If the state of one spin imposes no hard constraint on the state of the other, we may specify their states independently. For example, if they are far apart and prepared completely independently, we would expect that the probabilities for outcomes of measurements on 1 and 2 separately should be uncorrelated: $P(1,2) = P(1)P(2)$. We can achieve this if the state of the combined system is of the form $|a\rangle_1 \otimes |b\rangle_2$ where the inner product behaves as

$$(\langle c|_1 \otimes \langle d|_2)\,(|a\rangle_1 \otimes |b\rangle_2) = \langle c|a\rangle_1 \langle d|b\rangle_2 \ .$$

But now Axiom 1 tells us that we must allow superpositions of vectors to also describe allowed states. The resulting combined vector space is the *tensor product* of the two Hilbert spaces[22]:

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$$

a basis for which is

$$\{|\uparrow\rangle_1 \otimes |\uparrow\rangle_2, |\uparrow\rangle_1 \otimes |\downarrow\rangle_2, |\downarrow\rangle_1 \otimes |\uparrow\rangle_2, |\downarrow\rangle_1 \otimes |\downarrow\rangle_2 \ \} \tag{16}$$

or more succinctly:

$$\mathcal{H} = \text{span}\{|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle\}.$$

You have already encountered the tensor product in describing the Hilbert space of a particle moving in more than one dimension: to indicate a position basis vector we must specify both its x-position and its y-position.

Notice that generic vector in a tensor product $\mathcal{H}$ cannot be written as a product

$$|w\rangle = \sum_{i,m} w_{im} |i\rangle_1 \otimes |m\rangle_2 \neq |v^1\rangle_1 \otimes |v^2\rangle_2$$

[22] Note that Le Bellac regards the statement that composite systems are made via tensor product as an extra postulate. I think the above line of reasoning circumvents this.

for any $v^{1,2}$. This is only possible if the coefficient matrix factorizes as $w_{i,m} = v^1_i v^2_m$. A matrix that can be written this way has rank 1 – only a one-dimensional eigenspace of nonzero eigenvalues. If $|w\rangle$ cannot be written this way, the two subsystems in the state $|w\rangle$ are said to be *entangled*. The rank of the matrix w is called the *Schmidt number* of the state $|w\rangle$; $|w\rangle$ is entangled if the Schmidt number is bigger than 1.

Note: this operation of taking the tensor product of vector spaces should be distinguished from the direct *sum*. A rough but useful way to remember the distinction is: To specify a state in the tensor product of $\mathcal{H}_a$ and $\mathcal{H}_b$ means we must specify the state of *both a and b*. To specify a state in the direct sum we must specify whether the vector is a state of *a or b*. More precisely, if $\mathcal{H}_a = \text{span}\{|i\rangle, i = 1..N\}$ and $\mathcal{H}_b = \text{span}\{|r\rangle, r = 1..M\}$ then

$$\mathcal{H}_a \otimes \mathcal{H}_b = \text{span}\{|i\rangle \otimes |r\rangle, i = 1..N, r = 1..M\},$$

– the tensor product of an N-dimensional $\mathcal{H}$ and an M-dimensional one is NM-dimensional. In contrast,

$$\mathcal{H}_a \oplus \mathcal{H}_b = \text{span}\{|i\rangle, |r\rangle, i = 1..N, r = 1..M\}$$

is only $N + M$ dimensional.

A few simple comments about tensor products.

- Scalars can be moved between the tensor factors – the state $(z|a\rangle) \otimes |b\rangle = |a\rangle \otimes (z|b\rangle)$.
- If $|a_1\rangle$ is orthogonal to $|a_2\rangle$ in $\mathcal{H}_1$ then $|a_1\rangle \otimes |b_1\rangle \perp |a_2\rangle \otimes |b_2\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$.

We may now allow the spins to interact with each other by introducing a Hamlitonian which acts on $\mathcal{H}$. So we have to think about operators acting on the tensor product. In general such an operator is a sum of operators of the form $\mathbf{A}_1 \otimes \mathbf{A}_2$ where $\mathbf{A}_{1,2}$ act on subsystems $1, 2$.

A little more detail about matrix representation of tensor products

A few simple examples to illustrate the tensor product.

The Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is four dimensional; a vector in it has four components. In the basis (16), we have the representation

$$|\uparrow\rangle_1 \otimes |\uparrow\rangle_2 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \quad |\uparrow\rangle_1 \otimes |\downarrow\rangle_2 = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \quad |\downarrow\rangle_1 \otimes |\uparrow\rangle_2 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \quad |\downarrow\rangle_1 \otimes |\downarrow\rangle_2 = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

Note that the order is a convention. In this basis, then, operators look like (for example):

$$\mathbb{1}_1 \otimes \mathbb{1}_2 = \begin{pmatrix}\mathbb{1}_2 & 0\\0 & \mathbb{1}_2\end{pmatrix} = \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}, \quad \mathbb{1}_1 \otimes \boldsymbol{\sigma}_2^x = \begin{pmatrix}\boldsymbol{\sigma}_2^x & 0\\0 & \boldsymbol{\sigma}_2^x\end{pmatrix} = \begin{pmatrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix},$$

$$\boldsymbol{\sigma}_1^x \otimes \mathbb{1}_2 = \begin{pmatrix}0 & \mathbb{1}_2\\\mathbb{1}_2 & 0\end{pmatrix} = \begin{pmatrix}0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\end{pmatrix}, \quad \boldsymbol{\sigma}_1^x \otimes \boldsymbol{\sigma}_2^x = \begin{pmatrix}0 & \boldsymbol{\sigma}_2^x\\\boldsymbol{\sigma}_2^x & 0\end{pmatrix} = \begin{pmatrix}0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{pmatrix}.$$

Notice that if we re-ordered our basis as

$$\{|\uparrow\rangle_1 \otimes |\uparrow\rangle_2, |\downarrow\rangle_1 \otimes |\uparrow\rangle_2, |\uparrow\rangle_1 \otimes |\downarrow\rangle_2, |\downarrow\rangle_1 \otimes |\downarrow\rangle_2\}$$

it would interchange matrix representations of the second and third examples above.

1.7.2 Density Matrices aka Density Operators aka State Operators

Despite appearances, the step that we just made from one qbit to two qbits is significant. The reason for this is that if we don't pay attention to the second qbit, the first qbit will behave in a way which seems to violate the Axioms above. Those axioms are axioms for the quantum mechanical description of *everything*, the whole universe. We can't leave anything out. On the other hand, in practice, we always have to leave something out – we can't keep track of all the degrees of freedom to which our system is coupled (think: dust particles, stray photons, air molecules, the walls of the container). And in fact, if we limit our attention to part of a larger system the axioms are violated: states are not rays, evolution is not unitary, and measurements are not orthogonal projections. All these complications arise already for two qbits.

To begin, imagine our system consists of two qbits, one of which we'll call A and which we have in our hands and can measure as much as we want (consistent with QM, of course), and another, B, which is somehow obscure to us – it's locked in another room, maybe in another galaxy. We'd like to characterize what information about the system we can obtain in this circumstance. (The juicy bits of this discussion arise if A and B were able to interact at some point, so that the joint state is an entangled state, in the sense defined above. We can show that interactions between A and B were required to create an entangled state.)

First [following Preskill 2.3.1] consider the particular (entangled) state

$$|c\rangle_{AB} = a|\uparrow\rangle_A \otimes |\uparrow\rangle_B + b|\downarrow\rangle_A \otimes |\downarrow\rangle_B \ . \tag{17}$$

Suppose we measure the $\boldsymbol{\sigma}^z$ for qbit A:
with probability $|a|^2$ we get $|\uparrow\rangle$ and the composite system ends up in the state $|\uparrow\rangle_A \otimes |\uparrow\rangle_B$;
with probability $|b|^2 = 1 - |a|^2$, we get $|\downarrow\rangle$ and the composite system ends up in the state $|\downarrow\rangle_A \otimes |\downarrow\rangle_B$. In both cases, we end up with a definite state of B by measuring A. In this sense, the outcomes of successive measurements of $\boldsymbol{\sigma}^z_A$ and $\boldsymbol{\sigma}^z_B$ are perfectly *correlated*. (Note that the same would be true if $b = 0$, but that's less interesting.)

Suppose we are interested in making more general measurements of A (but still can't measure anything about B). The most general such observable is of the form $\mathbf{M}_A \otimes \mathbb{1}_B$ where $\mathbf{M}_A$ is a self-adjoint operator on $\mathcal{H}_A$. The expectation value of this observable in the particular state $|c\rangle$ above (using the fact that our basis is ON) is:

$$\begin{aligned} \langle c|\mathbf{M}_A \otimes \mathbb{1}_B|c\rangle &= \big(a^\star\langle\uparrow|_A\langle\uparrow|_B + b^\star\langle\downarrow|_A\langle\downarrow|_B\big)\mathbf{M}_A \otimes \mathbb{1}_B\,(a|\uparrow\rangle_A|\uparrow\rangle_B + b|\downarrow\rangle_A s|\downarrow\rangle_B) \\ &= |a|^2\langle\uparrow|\mathbf{M}_A|\uparrow\rangle_A + |b|^2\langle\downarrow|\mathbf{M}_A|\downarrow\rangle_A. \end{aligned} \tag{18}$$

In the last line, all mention of B has disappeared. We can rewrite this (and any other such calculation) as

$$\langle\mathbf{M}_A\rangle = \mathrm{tr}\,(\mathbf{M}_A\boldsymbol{\rho}_A) \tag{19}$$

where $\mathrm{tr}(\cdot) \equiv \sum_n \langle n|\cdot|n\rangle$ denotes the *trace* and

$$\boldsymbol{\rho}_A \equiv |a|^2|\uparrow\rangle\langle\uparrow| + |b|^2|\downarrow\rangle\langle\downarrow|$$

('rho') is called the *density operator* (or density matrix or state operator) for qbit A. Several questions need to answered now about $\boldsymbol{\rho}_A$:

What is the interpretation of the density operator?

We haven't had to specify anything about the observable $\mathbf{M}_A$ acting on A. Therefore: we can interpret $\boldsymbol{\rho}_A$ as an *ensemble* of quantum states, that is, a probability distribution on quantum states of A. We get the same result for $\langle\mathbf{M}_A\rangle$ (and anything else we might measure about A) if we say: with probability $p_\uparrow = |a|^2$, A is in the quantum state $|\uparrow\rangle$, with probability $p_\downarrow = |b|^2$, A is in the quantum state $|\downarrow\rangle$ (as we did above).

How did we construct $\boldsymbol{\rho}_A$ here?

We did it very explicitly for the particular state $|c\rangle$ in (18). More generally, putting the composite system in any state $|\psi\rangle_{AB}$, that procedure can be described as "tracing out B", by the following sequence of steps. Define the density matrix for a pure state $|\psi\rangle$ to be

$$\boldsymbol{\rho}_\psi = |\psi\rangle\langle\psi| \quad .$$

This is the density matrix if we don't forget anyone. In this case we can compute the expectation value of any operator on $\mathcal{H}$ by

$$\langle \mathbf{A} \rangle_\psi = \langle\psi|\mathbf{A}|\psi\rangle = \mathrm{tr}\,(\mathbf{A}\boldsymbol{\rho}_\psi)\,.$$

(Notice that this expression has the same structure as (19).

The reduced density matrix for tracing over a subsystem in the state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is then

$$\boldsymbol{\rho}_A = \mathrm{tr}_B \boldsymbol{\rho}_\psi = \mathrm{tr}_B\big(|\psi\rangle\langle\psi|\big).$$

We'll make this more explicit below in (20).

Note that the density operator is useful for emphasizing the distinction between quantum superposition and a mere probabilistic distribution, even for a single qbit. In the state $(|\uparrow_z\rangle + |\downarrow_z\rangle)/\sqrt{2} = |\uparrow_x\rangle$, the value of $\boldsymbol{\sigma}^z$ is uncertain, but the measurement of $\boldsymbol{\sigma}^x$ gives $+1$ with probability one. The ensemble in which $|\uparrow_z\rangle$ and $|\downarrow_z\rangle$ occur with probability $1/2$ is described by the density operator

$$\boldsymbol{\rho}_{\text{mixed}} \equiv \frac{1}{2}|\uparrow_z\rangle\langle\uparrow_z| + \frac{1}{2}|\downarrow_z\rangle\langle\downarrow_z| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \frac{1}{2}\mathbb{1}.$$

This is a very different thing from the (*pure*) state $|\uparrow_x\rangle$, which has a density operator

$$\boldsymbol{\rho}_{\text{pure}} = |\uparrow_x\rangle\langle\uparrow_x| = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\left(1/\sqrt{2}, 1/\sqrt{2}\right) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad .$$

Obviously these are different operators. To see how they give different predictions, consider the expectation value of the projection onto $|\uparrow_x\rangle$, $\mathbf{P}(\uparrow_x) = |\uparrow_x\rangle\langle\uparrow_x|$ in each of these two cases (this we interpret as the probability that we will get $\uparrow$ if we measure $\boldsymbol{\sigma}^x$). In the state $|\uparrow_x\rangle$, this is

$$\langle \mathbf{P}(\uparrow_x)\rangle_{\text{pure}} = \mathrm{tr}\boldsymbol{\rho}_{\text{pure}}\mathbf{P}(\uparrow_x) = \langle\uparrow_x|\uparrow_x\rangle\langle\uparrow_x|\uparrow_x\rangle = 1,$$

that is, like I said, we are certain to get $\uparrow$ if we measure $\boldsymbol{\sigma}^x$ in the state $|\uparrow_x\rangle$. On the other hand, in the ensemble $\boldsymbol{\rho}_{\text{mixed}}$, we get

$$\langle \mathbf{P}(\uparrow_x)\rangle_{\text{mixed}} = \mathrm{tr}\boldsymbol{\rho}_{\text{mixed}}\mathbf{P}(\uparrow_x) = \mathrm{tr}\frac{1}{2}\mathbb{1}|\uparrow_x\rangle\langle\uparrow_x| = \frac{1}{2}.$$

In fact, as you can see from the fact that $\boldsymbol{\rho}_{\text{mixed}}$ is proportional to the identity on $\mathcal{H}_A$, if we measure the spin along *any* axis in this state, we get a completely random result.

Now we turn to a general composite system (not necessarily two qbits). So far, we've focused on the particular entangled state $|c\rangle$ of two qbits in (17). The discussion of how to characterize our measurements of the subsystem A in terms of a density operator extend to any state of $\mathcal{H}_A \otimes \mathcal{H}_B$. Suppose that $\mathcal{H}_A = \text{span}\{|i\rangle_A\}, \mathcal{H}_B = \text{span}\{|r\rangle_B\}$ (maybe qbits, maybe not). then the most general state in $\mathcal{H}_A \otimes \mathcal{H}_B$ is

$$|a\rangle_{AB} = \sum_{ir} a_{ir}|i\rangle_A \otimes |r\rangle_B$$

which is normalized if $\sum_{ir} |a_{ir}|^2 = 1$. The expectation value of an observable acting only on A, $\mathbf{M}_A \otimes \mathbb{1}_B$ is

$$\begin{aligned}\langle \mathbf{M}_A \rangle &= \langle a|\mathbf{M}_A \otimes \mathbb{1}_B|a\rangle = \sum_{j,s}\sum_{i,r} a^\star_{js}\langle j|_A \otimes \langle s|_B \left(\mathbf{M}_A \otimes \mathbb{1}_B\right) a_{ir}|i\rangle_A \otimes |r\rangle_B \\ &= \sum_{ij,r} a_{ir}a^\star_{jr}\langle j|_A \mathbf{M}_A|i\rangle_A = \text{tr}_A \boldsymbol{\rho}_A \mathbf{M}_A\end{aligned}$$

with

$$\boldsymbol{\rho}_A = \sum_{ij,r} |i\rangle_{AA}\langle j| a_{ir}a^\star_{jr} \ . \tag{20}$$

The density operator $\boldsymbol{\rho}_A$ for subsystem A is obtained by performing a *partial trace* over subsystem B of the density matrix for the combined system. To get an expectation value, we take the rest of the trace.

Notice that if the state $|a\rangle$ is an unentangled state of $\mathcal{H}_A \otimes \mathcal{H}_B$, that is if $a_{ir} = v_i u_r$ has rank one, then $\boldsymbol{\rho}_A$ is a pure state:

$$\boldsymbol{\rho}_A = \sum_r u^\star_r u_r \sum_i v_i|i\rangle \sum_j v^\star_j\langle j| = |v\rangle\langle v|$$

with $|v\rangle \equiv \sum_i v_i|i\rangle \in \mathcal{H}_A$ (using the fact that the state is normalized). In this case, we can completely forget $\mathcal{H}_B$.

[End of Lecture 8]

In summary: The expectation value of an operator $\mathbf{A}$ acting on $\mathcal{H}$ in the state $\boldsymbol{\rho}$ (also acting on $\mathcal{H}$!) is

$$\langle \mathbf{A} \rangle_{\boldsymbol{\rho}} = \mathrm{tr}_{\mathcal{H}} \left(\boldsymbol{\rho} \mathbf{A} \right) \ .$$

The density matrix for a pure state $|\psi\rangle$ is

$$\boldsymbol{\rho}_\psi = |\psi\rangle\langle\psi| \ .$$

(Notice that the unphysical overall phase of $|\psi\rangle$ drops out of the density matrix.)

If we can make the decomposition $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ then the reduced density matrix

$$\boldsymbol{\rho}_A \equiv \mathrm{tr}_B \boldsymbol{\rho}_\psi$$

encodes all information about observables acting only on $\mathcal{H}_A$, *i.e.* of the form $\mathbf{A} \otimes \mathbb{1}_B$.

The following properties are satisfied by the expression (20) and are required for its interpretation as a probability distribution on quantum states:

What are the properties of the density operator?

In general the density operator

- has unit trace (because $|\psi\rangle$ is a normalized state, and it had better because its trace is the sum of all the probabilities), $1 = \langle \mathbb{1} \rangle = \mathrm{tr} \left(\boldsymbol{\rho}_A \right)$
- is self-adjoint, $\boldsymbol{\rho}_A = \boldsymbol{\rho}_A^\dagger$ (its eigenvalues are probabilities) and, moreover,
- is positive (all of its eigenvalues are positive or zero (and in particular real), since they are probabilities).

So: in general, the state of a subsystem is not a ray, it is a density operator. The density operator associated with a ray $|\psi\rangle$ (a pure state) is of the form $\boldsymbol{\rho}_{\mathrm{pure}} = |\psi\rangle\langle\psi|$, a projector of rank one, onto the space spanned by $|\psi\rangle$. $\boldsymbol{\rho}_{\mathrm{pure}}^2 = \boldsymbol{\rho}_{\mathrm{pure}}$. (Notice that although any projector satisfies this equation, it is only consistent with $\mathrm{tr}\boldsymbol{\rho} = 1$ if it is a projector of rank one.)

A more general ("mixed") density matrix can be written (by spectral decomposition – it is positive and hence Hermitian) as

$$\boldsymbol{\rho}_{\mathrm{mixed}} = \sum_a p_a |\psi_a\rangle\langle\psi_a|$$

with $0 \leq p_a \leq 1$, $\sum_a p_a = 1$. You should think of p_a as the probability that the system is found in the state $|\psi_a\rangle$ (an eigenstate of the density operator).

Density operator for a qbit

For the case of a single qbit, the most general density matrix is a hermitian 2×2 matrix, so can be written

$$\boldsymbol{\rho} = n_0 \mathbb{1} + \frac{1}{2} n_i \boldsymbol{\sigma}^i \ .$$

We must have $1 = \mathrm{tr}\boldsymbol{\rho} = 2n_0$ (recall that the $\boldsymbol{\sigma}^i$ are traceless), so

$$\boldsymbol{\rho} = \frac{1}{2}\left(\mathbb{1} + \vec{n}\cdot\vec{\boldsymbol{\sigma}}\right) = \frac{1}{2}\begin{pmatrix} 1+n_3 & n_1 - in_2 \\ n_1 + in_2 & 1-n_3 \end{pmatrix} \ . \tag{21}$$

Its determinant is $\det \boldsymbol{\rho} = \frac{1}{4}(1-\vec{n}^2)$. No negative eigenvalues requires $0 \leq \rho_1\rho_2 = \det\boldsymbol{\rho} \implies \vec{n}^2 \leq 1$. (This is in fact sufficient since $\mathrm{tr}\boldsymbol{\rho} = 1$ means at most one negative eigenvalue for the density matrix of a qbit.) So: possible density matrices of a qbit correspond to points inside the unit ball $\{\vec{n} | 0 \leq \vec{n}^2 \leq 1\}$. This is called the Bloch ball.

Its boundary $|\vec{n}| = 1$ is the Bloch sphere we studied earlier: it corresponds to density matrices whose determinant vanishes; since $\mathrm{tr}\boldsymbol{\rho} = 1$, the nonzero eigenvalue of such an operator must be 1, and these are projectors onto pure states. (We can see directly that $\boldsymbol{\rho}$ in (21) is a projector when $\vec{n}^2 = 1$ using $\vec{\boldsymbol{\sigma}}\cdot\hat{n}^2 = \mathbb{1}$.) With a little effort, we can figure out which pure state; you won't be too surprised that $\hat{n}$ is the direction in which the spin is pointing up.

Notice that while the ray description of the state $|\uparrow_{\hat{n}}\rangle$ involves a meaningless phase, its associated density matrix

$$\boldsymbol{\rho}_{\hat{n}} = \frac{1}{2}\left(\mathbb{1} + \hat{n}\cdot\vec{\boldsymbol{\sigma}}\right)$$

only involves physically meaningful quantities.

1.7.3 Time evolution of the density operator, first pass

Consider the time evolution for the density operator of a pure state:

$$\mathbf{P}_\psi(t) \equiv |\psi(t)\rangle\langle\psi(t)| \ .$$

Using our time-evolution axiom for states

$$\mathbf{i}\hbar\partial_t|\psi(t)\rangle = \mathbf{H}|\psi(t)\rangle$$

we infer that

$$\mathbf{i}\hbar\partial_t\mathbf{P}_\psi(t) = \mathbf{i}\hbar\partial_t\left(|\psi(t)\rangle\langle\psi(t)|\right) = \mathbf{H}\mathbf{P}_\psi - \mathbf{P}_\psi\mathbf{H} = [\mathbf{H}, \mathbf{P}_\psi].$$

This evolution is unitary; one way to explain this is that the finite time evolution (for t-independent $\mathbf{H}$) is:

$$\mathbf{P}(t) = e^{-\mathbf{i}\mathbf{H}t}\mathbf{P}(0)e^{+\mathbf{i}\mathbf{H}t}$$

and $e^{-\mathbf{i}\mathbf{H}t}$ is unitary because $\mathbf{H}$ is Hermitian.

Similarly, if our system is made of two subsystems which are *decoupled*, that is, if

$$\mathbf{H}_{AB} = \mathbf{H}_A \otimes \mathbb{1}_B + \mathbb{1}_A \otimes \mathbf{H}_B \tag{22}$$

then the evolution of the reduced density matrix for A is simple: this is because the second term in $\mathbf{H}_{AB}$ commutes with any operator acting only on A:

$$\mathbf{i}\hbar\partial_t \boldsymbol{\rho}_A = [\mathbf{H}_A \otimes \mathbb{1}_B + \mathbb{1}_A \otimes \mathbf{H}_B, \boldsymbol{\rho}_A \otimes \mathbb{1}_B] = [\mathbf{H}_A, \boldsymbol{\rho}_A] \quad .$$

This evolution is also unitary.

However, if the two systems *interact*, *i.e.* if the hamiltonian on the combined system is more general than (22), the evolution of $\boldsymbol{\rho}_A$ is not unitary: this is simply the statement that probability can leak from A into B, and vice versa. Making this quantitive is a more involved and will have to wait a bit.

1.8 Entanglement

[Preskill Chapter 2.4, Chapter 4; Le Bellac chapter 6; Weinberg chapter 12] Next we will begin to explore the consequences of entangled states.

1.8.1 "Spooky action at a distance"

Einstein Locality aka Local Realism

Reconsider the Stern-Gerlach experiment. Suppose we have a source that produces spins up along $\hat{z}$, and we use the device measure the spin along $\hat{x}$; quantum mechanics correctly tells us that we get two spots of equal intensity, but doesn't tell us the outcome of any given trial. You might think that our description of this process is correct, but *incomplete*: that is, perhaps there is some other data that the particle is carrying around with it that it uses to decide which of the two spots to hit. And maybe the probabilistic nature of QM arises because we are averaging over the values of these "hidden variables".

[End of Lecture 9]

For example, you could imagine a theory where the deeper description of a qbit in the pure state $|\uparrow_z\rangle$ is labelled by $(\hat{z}, \lambda)$, where the (unkown) value of $\lambda \in [0,1]$ (the hidden variable) decides whether we measure up or down when we measure some other spin direction θ, like:

$$\begin{aligned} |\uparrow_\theta\rangle, \qquad &\text{if} \quad 0 \le \lambda \le \cos^2\frac{\theta}{2} \\ |\downarrow_\theta\rangle, \qquad &\text{if} \quad \cos^2\frac{\theta}{2} \le \lambda \le 1 \ . \end{aligned} \tag{23}$$

The idea is that if we knew λ, all the quantum mechanical indeterminism would be gone. A *local* hidden variables theory is one in which, if A and B are out of causal contact, measurements of A do not affect the values of the hidden variables determining measurements of B. ('Out of causal contact' just means even light doesn't have enough time to get from A to B in time.) The question is whether a quantum measurement can be thought of as simply revealing some already-determined classical information stored in the relevant region of space-time.

We will see below that this possibility (which is called *Einstein locality* or *Local realism*) is ruled out by the observed violation of Bell's inequalities: if there are hidden variables, they aren't local; this probably means there aren't hidden variables in any useful sense.

Entangled states

The following observations were made by EPR [Einstein-Podolsky-Rosen] as evidence that quantum mechanics is incomplete if local realism is correct. Consider two particles (one dimension suffices) in a state which is an eigenstate of their relative position $x_1 - x_2$, where that eigenvalue R is large – *i.e.* they are far apart.

$$\psi(x_1, x_2) = \delta(x_1 - x_2 - R) = \frac{1}{2\pi} \int dk \; e^{ik(x_1 - x_2 - R)} \quad .$$

(Notice that this is state is also an eigenstate of the total momentum: $(\partial_{x_1} + \partial_{x_2})\,\psi = 0$.) In this state, if we measure the momentum of one particle to be k_1, we know for sure that the momentum of the other particle is $-k_1$, even though it might be at Alpha Centauri. The discomfort caused by that statement is the motivation for the discussion below. Such states can indeed be prepared, for example if the two particles are created in the decay of a bound state of the two, at rest.

A refinement of this situation (due to David Bohm) is to consider two particles with spin $\frac{1}{2}$. We are actually going to ignore their position degree of freedom, and just think about their spin.

Consider the state

$$|\text{Bohm}\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle \right) \equiv \frac{1}{\sqrt{2}} \left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle \right) \ . \tag{24}$$

(Sometimes called an EPR pair.) Notice that this state is a *singlet* of the total spin:

$$\vec{\boldsymbol{\sigma}}_T \equiv \vec{\boldsymbol{\sigma}}_A \otimes \mathbb{1}_B + \mathbb{1}_A \otimes \vec{\boldsymbol{\sigma}}_B \ , \tag{25}$$

(We will sometimes write things like this as $\boldsymbol{\sigma}_T = \boldsymbol{\sigma}_A + \boldsymbol{\sigma}_B$.) that is (this is three equations):

$$\vec{\boldsymbol{\sigma}}_T |\text{Bohm}\rangle = 0 \ . \tag{26}$$

This fact (that this is the state of two qbits with zero total angular momentum) can guarantee that when a spinless object decays into this pair of spin-1/2 particles, this is the spin state they end up in, and lets us imagine preparing this 2-qbit state.

In the state $|\text{Bohm}\rangle$, the two spins are *maximally entangled*. The meaning of this term is: if we compute the reduced density matrix for A, tracing over the states of B, we get an answer proportional to the identity:

$$\begin{aligned}\boldsymbol{\rho}_A &\equiv \text{tr}_B|\text{Bohm}\rangle\langle\text{Bohm}| \\ &= \sum_{\alpha_B=\uparrow,\downarrow} {}_B\langle\alpha_B|\frac{1}{\sqrt{2}}\left(|\uparrow\rangle_A\otimes|\downarrow\rangle_B - |\downarrow\rangle_A\otimes|\uparrow\rangle_B\right)\frac{1}{\sqrt{2}}\left(\langle\uparrow|_A\otimes\langle\downarrow|_B - \langle\downarrow|_A\otimes\langle\uparrow|_B\right)|\alpha_B\rangle_B \\ &= \frac{1}{2}\left(-|\downarrow\rangle_A\,(-{}_A\langle\downarrow|) + |\uparrow\rangle_A{}_A\langle\uparrow|\right) = \frac{1}{2}\mathbb{1}\end{aligned}$$

(the proportionality constant is fixed by $\text{tr}\boldsymbol{\rho} = 1$ to be $1/\dim\mathcal{H}_A$). Recall that this reduced density matrix for A encodes all the results of measurements that we can do if we only have access to A. This means that if we can't play with B also, we have absolutely no information about A!

Here is a nice way to quantify our ignorance about A given a density matrix for it. It is called the *von Neumann entropy*:

$$S(\boldsymbol{\rho}) \equiv -\text{tr}\boldsymbol{\rho}\log\boldsymbol{\rho}\ .$$

It eats a density matrix and gives a number. The log of a Hermitian operator can be defined in terms of the spectral decomposition:

$$\text{if}\ \ \boldsymbol{\rho} = \sum_a p_a|\psi_a\rangle\langle\psi_a|\ \ \text{then}\ \ \log(\boldsymbol{\rho}) = \sum_a \log(p_a)|\psi_a\rangle\langle\psi_a|\ .$$

Example 1 (certainty): If $\boldsymbol{\rho}_{\text{pure}} = |\psi\rangle\langle\psi|$ is a pure state in a Hilbert space of any dimension, then this is already a spectral decomposition and the only nonzero eigenvalue is $p_\psi = 1$, so

$$S(\boldsymbol{\rho}_{\text{pure}}) = -\text{tr}|\psi\rangle\langle\psi|\log(1) = 0.$$

In this case, we have zero ignorance about what quantum state we're in.

Example 2 (complete ignorance): If $\boldsymbol{\rho}_{\text{max}} = \frac{1}{N}\mathbb{1}_{N\times N}$ is the maximally-mixed density matrix on an N-dimensional Hilbert space (the reduced density matrix from tracing out B in a maximally-entangled state) then

$$S(\boldsymbol{\rho}_{\text{max}}) = -\text{tr}\left(\frac{1}{N}\mathbb{1}_{N\times N}\log\left(\frac{1}{N}\right)\right) = -\log\left(\frac{1}{N}\right)\cdot\frac{1}{N}\underbrace{\text{tr}\mathbb{1}_{N\times N}}_{=N} = +\log N.$$

I claim that this is the biggest answer you get get for the von Neumman entropy of an $N\times N$ density matrix. Maximal ignorance.

Yet another interpretation of eqn (26) is: it says that if we measure the spin of A and get σ, whichever axis we choose, if we measure the spin of B we will always get the opposite

answer $-\sigma$ for the spin of B along the same axis. In this sense, the two spins are perfectly anti-correlated.

So: if we made a list of the outcomes of Alice's results for measuring $\boldsymbol{\sigma}^{\hat{n}}$ it would look just like flipping a coin; but if we were able to compare Alice's list to Bob's list they would be perfectly anti-correlated.

Can we send information using this fact if the spins are far apart? Emphatically, no. No matter what measurement Bob does of his spin B, the density matrix for the spin A *is the same*! This means that no measurement we can do on A can ever tell anything about what measurement Bob did.

So the words in the title of this subsection (a quote from Einstein) are actually misleading: there is no *action* at a distance. There is still something weird about this situation, as we will see next.

Quantum erasure

This is a fancy name for the following fact: entanglement plus classical information can be used to make quantum information. Suppose we consider two qbits in the state

$$|c\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|\uparrow_z\rangle_A \otimes |\downarrow_z\rangle_B - |\downarrow_z\rangle_A \otimes |\uparrow_z\rangle_B \right) \ .$$

The reduced density matrix for A is totally random: $\boldsymbol{\rho}_A = \frac{1}{2}\mathbb{1}$.

If, however, we knew some classical information about B, like: a measurement of $\boldsymbol{\sigma}^z_B$ gives $\uparrow$, we would have to revise our information about A – we would know for sure that A is in the state $|\downarrow\rangle$, a pure state.

1.8.2 Bell's inequalities

[Le Bellac 6.3.2, Preskill 4.1.2, 4.1.3. See especially the discussion in the revised version of Preskill chapter 4.2.] As Le Bellac says well: the perfect (anti-)correlation of results when we measure $\boldsymbol{\sigma}^z$ is perfectly understandable to a classical entity: it's just like the particles agree before they start their journey to carry with them opposite spins. That's something that can happen classically[23]. But notice again that our equation above:

$$\vec{\boldsymbol{\sigma}}_T |\text{Bohm}\rangle = 0$$

[23] Two people meet in a room and each one puts one shoe from the same pair of shoes in his suitcase, randomly selected. Then they both travel a long time in opposite directions. Of course, if we meet one of them and he has the left shoe, then we learn for sure that the other one has the right shoe, even if he is very far away. No need for quantum mechanics so far.

says that we get opposite results *no matter which axis of spin we choose to measure*. Here the quantumness comes into it.

Bell's further innovation here was to consider measuring spins A and B along *different* axes from each other. This lets us consider a bunch of experiments we can do to the same state.

Following Preskill, here's the classical analog problem, where Bell's inequality is satisfied. Suppose we have flipped three coins (fair or not); they are lying on the table and each one has two possible outcomes (H or T), but we haven't looked at the results yet. Let's limit ourselves (for reasons described below) to looking at only two of them. We can look at any two.

Now, suppose that there are local hidden variables that give a complete description of this system. For flipping coins, this is the case: we can't follow the trajectories of the coins and so we average over this information and get a probability distribution. Then we can use ordinary probability theory to account for the outcomes of the measurements. No matter what the outcomes for the three coins are (HHH, HHT, HTT, TTT), two of the coins always have the same outcome. So if I let $P_{\text{same}}(i,j)$ be the probability that coins i,j have the same outcome (HH or TT) then we have:

$$P_{\text{same}}(1,2) + P_{\text{same}}(2,3) + P_{\text{same}}(3,1) \geq 1 \ , \tag{27}$$

no matter what probability distribution we pick for the coins. This is an example of a Bell inequality. It's something that's clearly true if local realism is a correct description of the world.

[End of Lecture 10]

Proof of (27). Let me be more explicit. If there are local hidden variables determining the results of the coin flips, then we can assign a probability to each set of outcomes of the coins $P(x,y,z) \geq 0$, $x,y,z = H$ or T, and this set of outcomes is exhaustive:

$$\sum x,y,z = H,T P(x,y,z) = 1.$$

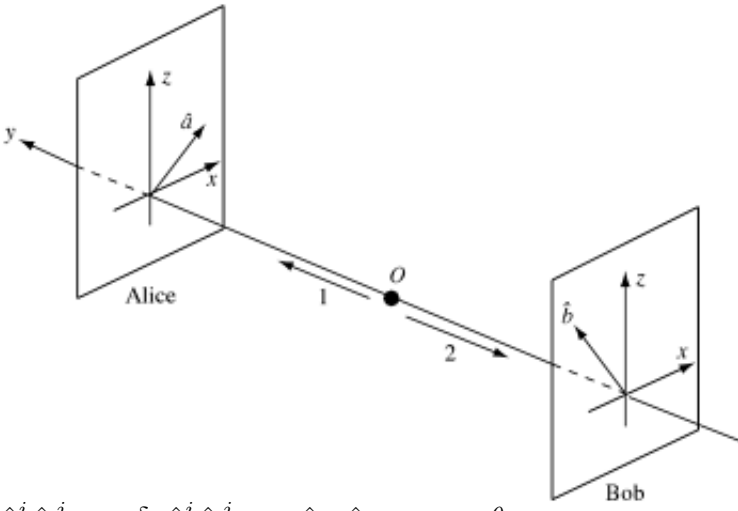
Then

$$\begin{aligned} P_{\text{same}}(1,2) &= P(HHH) + P(HHT) + P(TTH) + P(TTT) \\ P_{\text{same}}(2,3) &= P(HHH) + P(THH) + P(HTT) + P(TTT) \\ P_{\text{same}}(3,1) &= P(HHH) + P(HTH) + P(THT) + P(TTT) \end{aligned}$$

So the LHS of (27) is the sum of the LHSs of these equations. In the sum of the RHSs, $P(HHH)$ and $P(TTT)$ appear three times, and the other configurations appear once. So:

$$P_{\text{same}}(1,2) + P_{\text{same}}(2,3) + P_{\text{same}}(3,1) = \underbrace{\sum_{x,y,z=H,T} P(x,y,z)}_{} = 1 + 2(P(HHH) + P(TTT)) \geq 1$$

since each probability is positive.



The quantum experiment is to measure the spins A and B along the different axes $\hat{n}$ and $\hat{m}$. I will write $\vec{\boldsymbol{\sigma}}^{(A)} \equiv \vec{\boldsymbol{\sigma}}^{(A)} \otimes \mathbb{1}_B$. Consider the expectation value:

$$e(\hat{n},\hat{m}) = \langle \text{Bohm}| \left(\vec{\boldsymbol{\sigma}}^{(A)} \cdot \hat{n}\right) \left(\vec{\boldsymbol{\sigma}}^{(B)} \cdot \hat{m}\right) |\text{Bohm}\rangle$$

The singlet condition (26) means that when acting on $|\text{Bohm}\rangle$, we can replace $\vec{\boldsymbol{\sigma}}^{(A)}$ with $-\vec{\boldsymbol{\sigma}}^{(B)}$. So we can write this expectation value in terms of that of an operator on A:

$$e(\hat{n},\hat{m}) = -\langle \boldsymbol{\sigma}_i^{(A)} \boldsymbol{\sigma}_j^{(A)} \rangle \hat{n}^i \hat{m}^j = -\text{tr}\left(\boldsymbol{\rho}_A \boldsymbol{\sigma}_i^{(A)} \boldsymbol{\sigma}_j^{(A)}\right) \hat{n}^i \hat{m}^j = -\delta_{ij} \hat{n}^i \hat{m}^j = -\hat{n}\cdot\hat{m} \equiv -\cos\theta.$$

This agrees with our earlier statement that if we measure A and B along the same axis, we always get opposite answers.

The probability that A is up along $\hat{n}$ is the expectation value of the projection operator which projects onto that eigenstate of $\hat{n}\cdot\boldsymbol{\sigma}_A$, namely

$$\mathbf{E}_A(\hat{n},+) = \frac{1}{2}\left(\mathbb{1} + \hat{n}\cdot\boldsymbol{\sigma}^{(A)}\right) = |\uparrow_{\hat{n}}\rangle\langle\uparrow_{\hat{n}}|.$$

That is:

$$\textbf{Prob}\,(\text{A is}\uparrow_{\hat{n}}) = \text{tr}|\text{Bohm}\rangle\langle\text{Bohm}||\uparrow_{\hat{n}}\rangle\langle\uparrow_{\hat{n}}|$$

The probability that B is up along $\hat{m}$ is the expectation value of the projection operator $\mathbf{E}_B(\hat{n},+) = \frac{1}{2}\left(\mathbb{1} + \hat{m}\cdot\boldsymbol{\sigma}^{(B)}\right)$. The probability that A is up along $\hat{n}$ *and* B is up along $\hat{m}$ is the expectation value of product of these projection operators (they commute, so we can measure them simultaneously):

$$\textbf{Prob}\,(\text{A is}\uparrow_{\hat{n}}\text{ and B is}\uparrow_{\hat{m}}) = \langle \mathbf{E}_A(\hat{n},+)\mathbf{E}_B(\hat{m},+)\rangle = \frac{1}{4}\left(1-\cos\theta\right)\ .$$

By this kind of logic we learn that in the general case, the probabilities that A and B opposite or the same are

$$\textbf{Prob}(\text{opposite}) = \frac{1}{2}\left(1+\cos\theta\right), \quad \textbf{Prob}(\text{same}) = \frac{1}{2}\left(1-\cos\theta\right)\ .$$

Now let's consider three choices of $\hat{n}$ (designed to give simple answers for probabilities):

$$\hat{n}_1 = \hat{z}, \quad \hat{n}_2 = \left(\frac{\sqrt{3}}{2}, 0, -\frac{1}{2}\right), \quad \hat{n}_3 = \left(-\frac{\sqrt{3}}{2}, 0, -\frac{1}{2}\right)$$

(each one rotates by another 120°). If we measure $\sigma^{(A)}\cdot\hat{n}_1$, we put the system into the state $|\pm_{\hat{n}_1}\rangle_A \otimes |\mp_{\hat{n}_1}\rangle_B$, and we don't get to measure the second one. BUT: we can measure B

along $\hat{m}_1 \equiv -\hat{n}_2$, and we learn thereby what we *would* have gotten if we measured A along $\hat{n}_2$. The probability that the answers are the same is

$$\mathbf{Prob}_{\text{same}}(1,2) = \frac{1}{2}(1-\cos\theta) = \frac{1}{2}(1-\hat{n}_1\cdot(-\hat{n}_2)) = \frac{1}{2}\left(1-\frac{1}{2}\right) = \frac{1}{4}.$$

To be clear: this is the probability that we would get either $+,+$ or $-,-$ for the results of $\hat{n}_1\cdot\boldsymbol{\sigma}_A$ and $\hat{n}_2\cdot\boldsymbol{\sigma}_A$.

In the same way, by aligning the S-G apparatii for A and B, we can determine what we *would get*[24] if we measured the spin of A along any pair of $\hat{n}_1, \hat{n}_2, \hat{n}_3$. (This is why we consider only two coins at a time in the classical analog. It's like the third coin disappears when we look at two of them. Yes, that's already weird.)

SO: let's check whether Bell's inequality is satisfied by the quantum prediction:

$$\mathbf{Prob}_{\text{same}}(1,2) + \mathbf{Prob}_{\text{same}}(2,3) + \mathbf{Prob}_{\text{same}}(3,1) = 3\cdot\frac{1}{4} = \frac{3}{4} < 1.$$

What happened? We got in trouble by assigning probabilities to the outcomes of experiments that we didn't do, namely measuring the spin of A along two different axes. According to the rules of QM, we *cannot* do both measurements on the initial state. We *can* measure the spin of B, and the spin of A separately.

This violation has been seen experimentally [Clauser 1972 , Aspect 1982], using photons; see Preskill 4.1.4 for more details.

Comment: People are sometimes heard to say[25] that the observed violations of Bell's inequalities imply that "physics is non-local." This is very misleading! Rather, what they show is that *if we insist on a classical understanding of physics* (*i.e.* a "hidden variables" theory), that theory must be non-local. That is: it shows that *either* the world is non-local *or* it is quantum mechanical.

A nice mantra to keep in mind is:
"Unperformed experiments have no results." [Asher Peres]

The example above is a special case of the Bell inequality (and its violation by quantum mechanics and by the real world). This case can be generalized in many ways. We can consider other angles between the polarizers. We can consider more general entangled states than just the maximally-entangled Bohm state. In fact: any entangled pure state of two qbits violates some Bell inequality [Preskill 4.1.8]. (Not quite so for mixed states.)

[24] Beware the subjunctive in QM!

[25] *e.g.*: *How the Hippies Saved Physics*, by David Kaiser, pages 35 and 178, which has an otherwise-good discussion of Bell's inequalities and the experiments which observe their violation in our world.

1.8.3 GHZM states

[Le Bellac 6.3.4, Coleman's lecture] Something even more dramatic can happen with three qbits: a perfect anti-correlation between the correct answer from QM and the expectation from local realism.

Consider the following state of 3 qbits [GHZM stands for Greenberger, Horne, Zeilinger, Mermin]:

$$|GHZM\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\uparrow\uparrow\rangle - |\downarrow\downarrow\downarrow\rangle\right) \quad \in \mathcal{H} \equiv \mathcal{H}_a \otimes \mathcal{H}_b \otimes \mathcal{H}_c \ .$$

I am writing $|\uparrow\uparrow\uparrow\rangle \equiv |\uparrow\rangle_a \otimes |\uparrow\rangle_b \otimes |\uparrow\rangle_c$ to avoid clutter. We can imagine preparing three qbits in this state and sending them to distant laboratories, labelled a, b, c. The observers at these laboratories measure the spins along various axes. Let A_x denote the result of measuring the x-spin of particle a by the observer who catches it, B_y denote the result of measuring the y-spin of particle b, etc...; each of these satisfies $A_x^2 = 1$ so $A_x \pm 1$.

The classical situation to which we'd like to compare is if these measurements are all completely independent and don't care about each others' results. If all these measurements could be done independently (as in a classical world with local hidden variables), the outcomes of these results would obviously satisfy

$$A_x B_x C_x = (A_x B_y C_y)(A_y B_x C_y)(A_y B_y C_x) \tag{28}$$

(using $A_y^2 = B_y^2 = C_y^2 = 1$).

Now we do quantum mechanics instead. Consider the following operators acting on $\mathcal{H}_a \otimes \mathcal{H}_b \otimes \mathcal{H}_c$:

$$\boldsymbol{\Sigma}_a \equiv \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y, \quad \boldsymbol{\Sigma}_b \equiv \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_y, \quad \boldsymbol{\Sigma}_c \equiv \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_x \quad .$$

Now I am relying on the order to keep track of which qbit each sigma is acting on – the fully explicit expression is:

$$\boldsymbol{\Sigma}_a \equiv \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y \equiv \boldsymbol{\sigma}_{ax} \otimes \boldsymbol{\sigma}_{by} \otimes \boldsymbol{\sigma}_{cy}.$$

Notice that the label on the $\boldsymbol{\Sigma}_i$ indicates which of the three factors has the $\boldsymbol{\sigma}_x$. The operator

$$\boldsymbol{\Sigma} \equiv \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x$$

will also be useful.

Claims:

(a) These operators $\boldsymbol{\Sigma}_i$ all commute with each other. This can be checked using the properties of the Paulis – the important point is that $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_j$ differ by an even number of anti commutations. They also commute with $\boldsymbol{\Sigma}$.

(b) It is therefore possible for them to be simultaneously measured. Notice that this will require some coordinated effort between the observers in those distant laboratories.

(c) The $\mathbf{\Sigma}$s square to 1 and hence have eigenvalues ± 1.

(d) $\mathbf{\Sigma}_i|GHZM\rangle = |GHZM\rangle, i = a, b, c$. The state $|GHZM\rangle$ above is an eigenstate of all three $\mathbf{\Sigma}$s with eigenvalue $+1$. Act with $\mathbf{\Sigma}$ on it and see what happens! (Recall that $\boldsymbol{\sigma}_x|\uparrow\rangle = |\downarrow\rangle$...)

(e) $\mathbf{\Sigma}|GHZM\rangle = -|GHZM\rangle$ – it's also an eigenstate of $\mathbf{\Sigma}$, with eigenvalue -1.

Now comes the icepick to the forehead: If the observers measure one x component and two y components of the qbits in the GHZM state, the outcome is an eigenvalue of one of the operators $\mathbf{\Sigma}_i$, with eigenvalue $+1$:

$$A_x B_y C_y = +1, \quad A_y B_x C_y = +1, \quad A_y B_y C_y = +1 \ .$$

On the other hand, if two of the observers decide to rotate their polarizers so that they all measure in the x direction, they get the eigenvalue of $\mathbf{\Sigma}$:

$$A_x B_x C_x = -1 \ .$$

Now compare this with the classical expectation in (28)!

They disagree completely. The quantum answer is the correct one in our world.

Notice that although the Σs all commute with each other, they are made from operators which do not individually commute (*e.g.* $\boldsymbol{\sigma}_x$ and $\boldsymbol{\sigma}_y$ acting on any one of the qbits). We ran into trouble in the classical calculation because we assigned outcomes to these measurements of $\boldsymbol{\sigma}_x$ that we didn't do!

[End of Lecture 11]

We can use this discovery here to do something that you can't do without QM [I learned this from Boccio]:

Consider a game for three players A, B, C. They are told that they will be separated from each other and each one will be asked one of two questions, say X or Y whose allowed answers are $+1$ or -1. And *either*

(a) all players will be asked the same question X

or

(b) one of the three players will be asked X and the other two will be asked Y.

After having been asked X or Y no player can communicate with the others until all three players have given all their answers. To win, the players must give answers such that, in case (a) the produce of the three answers is -1, and in case (b) the product of the answers is $+1$.

What we've shown above is:

(1) There is no classical strategy that gives a certain win for the team.

(2) There is a quantum strategy which gives a certain win, as long as each player can take one of three qbits along with them, and the three qbits are in the GHZ state.

1.9 Things you can't do with classical mechanics

1.9.1 Uses of entanglement: quantum information processing

[Rieffel and Polak, *Quantum Computing, A Gentle Introduction*, chapter 5, Le Bellac 6.4.2] In deference to the subject matter, for this section only, I'm going to use computer-science notation. So I'll write

$$|\uparrow_z\rangle \equiv |0\rangle, \quad |\downarrow_z\rangle \equiv |1\rangle, \quad |\uparrow_x\rangle \equiv |+\rangle, \quad |\downarrow_x\rangle \equiv |-\rangle.$$

Also, this basis $\{|0\rangle, |1\rangle\}$ is called the 'computational basis' and is preferred – you should imagine that the operators we can measure with our quantum computer are diagonal in this basis.

Quantum gates

Like in our discussion of interferometers, we're going to think about subjecting our qbits to various unitary operators. The reason we focus on unitary operators is because we imagine (quite reasonably) that our quantum computer is governed by the laws of quantum mechanics, and its operation proceeds via unitary time evolution $\mathbf{U} = e^{i\mathbf{H}t}$; what we are doing is choosing the hamiltonian $\mathbf{H}$.

Some of these unitaries have traditional names.

$$\mathbf{X} \equiv \boldsymbol{\sigma}^x, \mathbf{Z} \equiv \boldsymbol{\sigma}^z, i\mathbf{Y} \equiv \sigma^y$$

are unitary operators.

Another important unitary acting on a single qbit is:

$$\mathbf{H} \equiv \frac{1}{\sqrt{2}} \left(|0\rangle\langle 0| - |1\rangle\langle 1| + |0\rangle\langle 1| + |1\rangle\langle 0|\right) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

(the last expression is the matrix representation in the computational basis) is called a *Hadamard gate*. It's the thing that a half-silvered mirror does. It's unitary and $\mathbf{H}^2 = \mathbb{1}$. Its purpose in life is to make (uniform) superpositions out of the computational basis states, and vice versa.

This next example of an often-arising gate acts on a pair of qbits:

$$\mathcal{C}_{\text{not}} \equiv |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes \mathbf{X}$$

is called the *controlled not* (or cnot) gate. Heuristically it does the following: if the first qbit is up, it does nothing; if the first qbit is down, it flips the second qbit. 'Flips the second qbit' is flippant language for 'acts by $\mathbf{X} \equiv \boldsymbol{\sigma}^x$'. It's also unitary and $\mathcal{C}_{\text{not}}^2 = \mathbb{1}$. The point

about cnot is that it's not of the form $\mathbf{A} \otimes \mathbf{B}$ – it changes the entanglement between the two qbits, for example it takes the product state

$$\mathcal{C}_{\rm not}|+\rangle \otimes |0\rangle = \frac{1}{\sqrt{2}}\mathcal{C}_{\rm not}\left(|00\rangle + |10\rangle\right) = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)$$

which is (maximally) entangled.

'No quantum xerox' or 'No cloning theorem'

Here's something you *can't* do with quantum mechanics: using unitary evolution, you cannot make a copy of an unknown quantum state. Such a map would take

$$\mathbf{Xerox} : |a\rangle \otimes |\text{anything}\rangle \to |a\rangle \otimes |a\rangle \ .$$

Similarly,

$$\mathbf{Xerox}|b\rangle \otimes |\text{anything}\rangle = |b\rangle \otimes |b\rangle \ .$$

But then what does it do to the superposition?

$$\mathbf{Xerox}\left(\left(|a\rangle + |b\rangle\right)/\sqrt{2}\right) \otimes |\text{anything}\rangle = \left(|a\rangle + |b\rangle\right)/\sqrt{2} \otimes \left(|a\rangle + |b\rangle\right)/\sqrt{2} \ .$$

But that 's not the same as the superposition of the images:

$$\mathbf{Xerox}\left(\left(|a\rangle + |b\rangle\right)/\sqrt{2} \otimes |x\rangle\right) \neq \frac{1}{\sqrt{2}}\left(|a\rangle \otimes |a\rangle + |b\rangle \otimes |b\rangle\right) = \frac{1}{\sqrt{2}}\left(\mathbf{Xerox}|a\rangle \otimes |x\rangle + \mathbf{Xerox}|b\rangle \otimes |x\rangle\right).$$

So such a map as **Xerox** can't even be linear, never mind unitary.

You can find operators that copy specific known states, but never arbitrary superpositions.

Quantum teleportation

Here's a work-around for moving around quantum information: We can transmit the unknown quantum state of a qbit by sending two classical bits, using some entanglement. The initial qbit state is destroyed in the process, in a way consistent with the no-cloning theorem.

What these words mean is the following. Suppose Alice and Bob each have one part of an EPR pair.

$$|\psi_0\rangle \equiv \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right) \in \mathcal{H}_a \otimes \mathcal{H}_b.$$

And suppose Alice also has in her hands another qbit in the state of interest

$$|\phi\rangle = a|0\rangle + b|1\rangle \in \mathcal{H}_c \ .$$

So the whole state is initially

$$\mathcal{H}_c \otimes \mathcal{H}_a \otimes \mathcal{H}_b \ni |\phi\rangle \otimes |\psi_0\rangle \equiv |\Psi_1\rangle \ .$$

And we are imagining that Alice controls c and a, while Bob, elsewhere, controls b.

Here's the protocol: Alice acts on the state by $\mathcal{C}_{\rm not}$ followed by the Hadamard $\mathbf{H}$ on the qbit to be teleported:

$$|\Psi_2\rangle = (\mathbf{H} \otimes \mathbb{1} \otimes \mathbb{1})\,(\mathcal{C}_{\rm not} \otimes \mathbb{1})\,|\Psi_1\rangle$$

The result is:

$$|\Psi_2\rangle = \frac{1}{2}\left(|00\rangle\,(a|0\rangle + b|1\rangle) + |11\rangle\,(a|1\rangle - b|0\rangle) + |10\rangle\,(a|0\rangle - b|1\rangle) + |01\rangle\,(a|1\rangle + b|0\rangle)\right).$$

Then Alice measures the z-spin of her two qbits, *i.e.* $\mathbf{Z} \equiv \boldsymbol{\sigma}^3$. The outcomes are pairs of $\pm$. Let's translate this to binary: instead let Alice measure

$$\frac{1}{2}\,(\mathbb{1} - \boldsymbol{\sigma}^z) = |1\rangle\langle 1|$$

(which gives either 0 or 1) for each of the two qbits. The outcomes comprise a pair of binary digits, that is, two bits. She then sends these two bits to Bob.

Now Bob follows the instructions in the table 1. If he receives 00, he doesn't do anything.

Bits sent	state of Bob's qbit	Bob's decoding operation
00	$a\|0\rangle + b\|1\rangle$	$\mathbb{1}$
01	$a\|1\rangle + b\|0\rangle$	$\mathbf{X}$
10	$a\|0\rangle - b\|1\rangle$	$\mathbf{Z}$
11	$a\|1\rangle - b\|0\rangle$	$\mathbf{Y}$

Table 1: Instructions for Bob.

The state of his qbit is now the input state. If he receives 01, he acts with $\mathbf{X} \equiv \boldsymbol{\sigma}^x$; again, the state of his qbit is now exactly the input state $|\phi\rangle$! Same for the other three cases.

Perhaps you should be surprised that we can send a whole Bloch-sphere's worth of quantum information using two measly bits. But don't forget the entanglement. This another manifestation of my earlier statement that classical info plus entanglement can be equivalent to quantum info.

Comments:

- Notice that the quantum information people are really good at giving their stuff cool-sounding names. (It goes a long way.) This is pretty far from Star Trek.
- Neither Alice nor Bob ever learns the values of the complex numbers a and b in the state $|\phi\rangle$. Alice's copy of the state is destroyed in this process, so this procedure is consistent with unitary evolution and the no-cloning theorem.

- Bob has to wait for the classical bits to reach him via some ordinary causal propagation of information in order to construct the state $|\phi\rangle$. Causality is OK.
- It is also worth mentioning that people have actually done this quantum teleportation business (references are in footnote 35 on p. 195 of Le Bellac).

Quantum dense coding

Quantum dense coding achieves the inverse of the previous goal, by doing literally the inverse operation. It uses entanglement to send two classical bits between Alice and Bob, by physically sending only one quantum bit. Now we imagine that Alice wants to send two classical bits to Bob – this means she wants to send a number from 0 to 3, or, in binary, one of 00, 01, 10, 11.

[End of Lecture 12]

We imagine again that they share an EPR pair of qbits in the state

$$|\psi_0\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B \right) \ .$$

Alice can only act on the first qbit and Bob can only act on the second.

Alice follows the protocol in Table 2. Then Alice sends her qbit to Bob.

Value to send	Alice acts with	resulting state $\|\psi_1\rangle$
$0 = 00$	$\mathbb{1} \otimes \mathbb{1}$	$\frac{1}{\sqrt{2}} \left(\|0\rangle_A \otimes \|0\rangle_B + \|1\rangle_A \otimes \|1\rangle_B \right)$
$1 = 01$	$\mathbf{X} \otimes \mathbb{1}$	$\frac{1}{\sqrt{2}} \left(\|1\rangle_A \otimes \|0\rangle_B + \|0\rangle_A \otimes \|1\rangle_B \right)$
$2 = 10$	$\mathbf{Z} \otimes \mathbb{1}$	$\frac{1}{\sqrt{2}} \left(\|0\rangle_A \otimes \|0\rangle_B - \|1\rangle_A \otimes \|1\rangle_B \right)$
$3 = 11$	$\mathbf{Y} \otimes \mathbb{1}$	$\frac{1}{\sqrt{2}} \left(-\|1\rangle_A \otimes \|0\rangle_B + \|0\rangle_A \otimes \|1\rangle_B \right)$

Table 2: Instructions for Alice's encoding.

Once he receives the qbit from Alice, Bob's decoding operation is to act on the state of the pair of qbits $|\psi_1\rangle$ by

$$(\mathbf{H} \otimes \mathbb{1}) \, \mathcal{C}_{\text{not}} |\psi_1\rangle$$

The first $\mathcal{C}_{\text{not}}$ operation unentangles the two qbits, and the Hadamard gate removes the superpositions in the 'computational basis', with the results described in 3. Then Bob measures $-\frac{1}{2}\boldsymbol{\sigma}^z + \frac{1}{2}\mathbb{1} = |1\rangle\langle 1|$ for each of the two qbits (these two observables commute); the results are exactly a binary representation of Alice's two bits.

$\lvert\psi_1\rangle$	$\mathcal{C}_{\text{not}}\lvert\psi_1\rangle$	$(\mathbf{H}\otimes\mathbb{1})\,\mathcal{C}_{\text{not}}\lvert\psi_1\rangle$
$\frac{1}{\sqrt{2}}(\lvert 00\rangle + \lvert 11\rangle)$	$\frac{1}{\sqrt{2}}(\lvert 00\rangle + \lvert 10\rangle) = \frac{1}{\sqrt{2}}(\lvert 0\rangle + \lvert 1\rangle)\otimes\lvert 0\rangle$	$\lvert 00\rangle$
$\frac{1}{\sqrt{2}}(\lvert 10\rangle + \lvert 01\rangle)$	$\frac{1}{\sqrt{2}}(\lvert 11\rangle + \lvert 01\rangle) = \frac{1}{\sqrt{2}}(\lvert 1\rangle + \lvert 0\rangle)\otimes\lvert 1\rangle$	$\lvert 01\rangle$
$\frac{1}{\sqrt{2}}(\lvert 00\rangle - \lvert 11\rangle)$	$\frac{1}{\sqrt{2}}(\lvert 00\rangle - \lvert 10\rangle) = \frac{1}{\sqrt{2}}(\lvert 0\rangle - \lvert 1\rangle)\otimes\lvert 0\rangle$	$\lvert 10\rangle$
$\frac{1}{\sqrt{2}}(-\lvert 10\rangle + \lvert 01\rangle)$	$\frac{1}{\sqrt{2}}(-\lvert 11\rangle + \lvert 01\rangle) = \frac{1}{\sqrt{2}}(-\lvert 1\rangle + \lvert 0\rangle)\otimes\lvert 1\rangle$	$\lvert 11\rangle$

Table 3: Results of Bob's decoding.

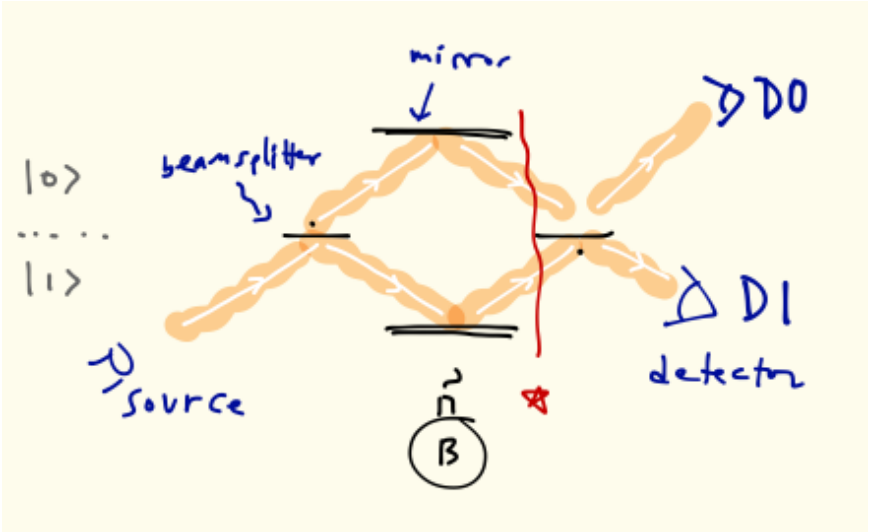
1.9.2 Exploiting quantum information

The magic tricks in this subsection don't really exploit entanglement as in the previous examples. The seeming magic is all in the incompatibilities of various measurements.

Quantum bomb testing

[Schumacher Chapter 1] Suppose you have a collection of bombs, and some of them may be faulty. You want a way to test them reliably which doesn't involve blowing up the ones that work. Here is a way to do this using quantum mechanics [due to Elitzur and Vaidman, 1993].

Consider the arrangement of beam-splitters and mirrors in the figure:



We denote the photon on the path above the dotted line by the state $|0\rangle$ and the photon on the path below the dotted line by the state $|1\rangle$.

Recall from section 1.6.4 that the beamsplitters can be realized by half-silvered mirrors.

Recall that these perform the unitary operation $\mathbf{H}$, the Hadamard gate. There is this annoying dot which indicates where the $-$ goes, so there are really two such gates:

$$\mathbf{H}_1|0\rangle = |+\rangle, \mathbf{H}_1|1\rangle = |-\rangle, \mathbf{H}_1|+\rangle = |0\rangle, \mathbf{H}_1|-\rangle = |1\rangle \quad .$$

$$\mathbf{H}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

The other one is:

$$\mathbf{H}_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

In the figure the left one is $\mathbf{H}_0$ and the right one is $\mathbf{H}_1$.

Claim 1: without any mirror at B, there is a signal at the detector $D1$ (the thing that looks like an eyeball).

Denote the top and bottom paths by the states $|0\rangle$ and $|1\rangle$ respectively. The input beam at the far right is in state $|1\rangle$, the lower path. Without the mirror at B, the state of the photon at $\star$ is just $|0\rangle$ – for sure it takes the upper path. Any photons that take the lower path would be lost, so if we see a photon it must have taken the upper path. The action of the second half-silvered mirror on this state $|1\rangle$ is:

$$\mathbf{H}_1|0\rangle = |+\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle\right).$$

So in this case we have probability $\left(\frac{1}{\sqrt{2}}\right)^2 = \frac{1}{2}$ of detecting a photon in the detector on the lower path, D1. [26]

Claim 2: When we put the mirror at B, the destructive interference between the two beams cancels the signal. We detect no photons at D1. To see this more explicitly, in this case the state of the photon at checkpoint $\star$ (we're not actually putting a detector there, though!) is a uniform superposition of the upper and lower paths: $|+\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle\right) = \mathbf{H}_0|0\rangle$, and so the action of the final beamsplitter is:

$$\mathbf{H}_1|+\rangle = |0\rangle;$$

for sure the photon goes to the upper path after passing the final beamsplitter. In the computational basis, we used

$$\mathbf{H}_0\mathbf{H}_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Now we are ready to use this device to make a non-destructive bomb-tester. Replace the mirror at B with the following kind of bomb detonator: if the bomb is a live, working bomb,

[26] Note that I am omitting the sign change caused by the (fully-silvered) mirrors; it just multiplies the whole wavefunction by -1 and so doesn't change any physics in this problem.

a photon striking it will cause the bomb to explode and be absorbed. On the other hand, if the bomb is a dud, the photon is reflected. This is something you could imagine engineering.

So: with this new device at B, what is the state of the photon at $\star$? The joint state of the photon and the bomb is :

$$\frac{1}{\sqrt{2}}\left(|+\rangle \otimes |\text{unexploded}\rangle + |0\rangle \otimes |\text{exploded}\rangle\right) \quad .$$

Why is this?: if the bomb did not explode, then photon is in the state resulting from a mirror, namely $|+\rangle$. If the bomb exploded, it's like no mirror: $|0\rangle$.

Now this state passes through the final beamsplitter like before, and we get the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \otimes |\text{unexploded}\rangle + |+\rangle \otimes |\text{exploded}\rangle\right) \quad .$$

What do we see at the detector? The claim is that any detection of a photon results from a live bomb that *did not explode*.

When I first heard about this protocol, I was skeptical about it. I thought that the entanglement of the photon state with the state of bomb, a macroscopic object, would cause problems for the delicate interference required to get the result. But my concerns were unfounded, as we can see by thinking about what happens if we explicitly include and then trace over the Hilbert space of the bomb in our calculation, as follows.

We should trace over the state of the bomb because the detector can be put arbitrarily far away. We are *not* measuring the operator $|\text{exploded}\rangle\langle\text{exploded}|$. The final density matrix for the photon (after it passes the last beamsplitter) is

$$\begin{aligned}
\boldsymbol{\rho} = \text{tr}_{\mathcal{H}_{\text{bomb}}}|\psi\rangle\langle\psi| = \text{tr}_{\mathcal{H}_{\text{bomb}}} \quad & \left(\frac{1}{\sqrt{2}}\left(|0\rangle \otimes |\text{unexploded}\rangle + |+\rangle \otimes |\text{exploded}\rangle\right)\right. \\
& \left.\frac{1}{\sqrt{2}}\left(\langle 0| \otimes \langle\text{unexploded}| + \langle + | \otimes \langle\text{exploded}|\right)\right) \\
& = \frac{1}{2}\left(|0\rangle\langle 0| + |+\rangle\langle + |\right) \\
& = \frac{1}{2}\left(|0\rangle\langle 0| + \frac{1}{2}\left(|0\rangle + |1\rangle\right)\left(\langle 0| + \langle 1|\right)\right) \\
& = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}\left(|0\rangle\langle 1| + |1\rangle\langle 0|\right) + \frac{1}{4}|1\rangle\langle 1| = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \quad .
\end{aligned}$$

[27] We conclude from this that the expected fraction of photons at the detector is

$$\langle \mathbf{P}_1 \rangle = \text{tr}\left(\boldsymbol{\rho}|1\rangle\langle 1|\right) = \frac{1}{4} \; !$$

But the only way a photon gets through is if a photon is absorbed by an exploding bomb!

[27] Note that this is a perfectly cromulent density matrix: its trace is 1, and its eigenvalues are $\frac{1}{2} \pm \frac{1}{\sqrt{2}} > 0$.

	Bomb is working	Bomb is a dud
Prob(photon reaches D1)	$\frac{1}{4}$	0
Prob(bomb explodes)	$\frac{1}{2}$	0
Prob(photon reaches D0)	$\frac{1}{4}$	1

Table 4: Probabilities for the two cases. Note that the columns each add up to one.

So: if the bomb explodes, it worked but we can't use it again. If the photon is detected at D0, the upper detector, the test is inconclusive. But: if ever a photon is detected at D1, we know the bomb was working! There is zero probability for this outcome if the bomb is a dud. Notice that the bomb did not detect the photon bouncing off its mirror; if it had it would have exploded. Trippy.

Quantum cryptography (quantum key distribution)

[Le Bellac 3.1.3]

This protocol uses quantum mechanics to ensure that a secret message was not spied upon. You should read the relevant section of Le Bellac.

1.9.3 Quantum algorithms

It might be possible to build a *quantum computer*. As Preskill says, your laptop is a computer whose operation involves quantum mechanics (lots of solid state physics), but it is not a quantum computer. By a quantum computer we mean an assembly of quantum systems (*e.g.*qbits) that we can manipulate through unitary gates and make measurements on. If we had such a device what could we do that we couldn't do otherwise? Even if it turns out that their construction is not feasible for another fifty years, this has been a fruitful question for learning about what is essential in quantum mechanics.

Actually, there is no computation we can do on a quantum computer that we couldn't do on a classical computer, if we are willing to wait long enough and make the computer big enough. That is, any model of what a quantum computer might do can be *simulated* by a (classical) Turing machine, and some patience.

But there are some things we can do faster. The useful way computer scientists compare these things is by how the required number of operations ('flops') of the computer depends on the size of the problem, as the size of the problem grows. For example, suppose you want to find the prime factors of an N-digit number n (so $N \sim \log(n)$). As N gets bigger, it takes longer and longer to reliably do this. The security of a lot of banking nonsense that goes on on the internet relies on the fact that for large enough N, this takes prohibitively long. The best known classical algorithm takes a time of order $e^{N^{1/3}}$.

With a quantum computer, you could solve this problem in a time polynomial in N. We will not discuss Shor's algorithm here (the number theory involved is too much of a distraction), but we will discuss a different, also very important problem whose solution can be sped up using quantum mechanics.

[End of Lecture 13]

Grover's algorithm

Looking for things classically is annoying. Suppose you want to find a thing that can be in any one of N boxes. Basically, you have to look in all the boxes until you find it. If the probability is uniform over the boxes, on average it takes $N/2$ looks before you find the thing. Quantumly, you can do better: of order $\sqrt{N}$ steps.

Perhaps looking for something you put in a box is a not vivid enough example. Instead, you can imagine any problem where you can guess the answer and quickly check that it's right. (The class of such problems is called NP.) A not-great example is finding the prime factors of a large integer. (It's not great because it hasn't been proved to be hard, and we already know an efficient quantum algorithm.) A better example is finding a Hamiltonian cycle in a graph of N nodes (a path through the graph that visits each node exactly once) or the problem of deciding whether there exists a route for a traveling salesman with N stops better than a given one. The latter two problems have the property that someone (Cook and Karp) proved that if you can solve these efficiently, you can efficiently solve any other problem in NP.

We divide this discussion in two parts. In the first part we show that given the ability to perform two unitary operations on an N-state hilbert space, one of which is the implementation of looking in the box, we can find the box with high probability.

In the second part of the discussion, we talk about the ability to do these operations given some basic building blocks (the unitary gates discussed above).

We want to find a, where $a \in \{1, 2...N\}$. How do we know which is a? We know it when we see it. What this means is that we have an *oracle*, whom we ask, 'is this a'? The oracle says 'yes' or 'no' and is always right.

Quantumly, let's suppose we can associate each elements of the set with an element of an orthonormal basis set for an N-state Hilbert space, $\mathcal{H} \equiv \text{span}\{|1\rangle, |2\rangle, ...|N\rangle\}$. This is called the 'computational basis' and is preferred – you should imagine that the operators we can measure with our quantum computer are diagonal in this basis. The quantum oracle is a Hermitian operator for which $|a\rangle$ is an eigenstate of eigenvalue -1 ('yes') and any orthogonal state is an eigenstate of eigenvalue $+1$ ('no'). This operator is

$$\mathcal{O} \equiv \mathbb{1} - 2|a\rangle\langle a|.$$

('$\mathcal{O}$' is for oracle.)

We initialize our computer to the uniform superposition of the computational basis states:

$$|u\rangle \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} |i\rangle \ .$$

In this state, the probability that we get the right answer if we make our measurement is $\cos^2\theta = \frac{1}{N}$ with $\cos\theta \equiv |\langle a|u\rangle| = \frac{1}{\sqrt{N}}$. I'll also use the complementary angle $\alpha = \pi/2 - \theta$ with $\sin\alpha = \cos\theta$. (See the figures below.)

We'll need a second operator, made from the projector onto $|u\rangle$:

$$\mathbf{P}_u \equiv |u\rangle\langle u|.$$

In our basis, this thing is the matrix with $1/N$ in every entry. It has rank 1. The operator we need is:

$$\mathbf{W} \equiv 2\mathbf{P}_u - \mathbb{1}.$$

It does nothing to $|u\rangle$, but flips the sign of any orthogonal vector.

Grover's idea is to do stuff repeatedly to the state to improve our chances of getting it right. The thing we repeat is $\mathbf{W}\mathcal{O}$.

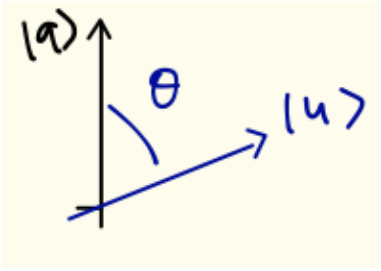
Idea: Start in the state $|u\rangle$:

$$\mathbf{W}\mathcal{O}|u\rangle = \mathbf{W}\frac{1}{\sqrt{N}}\left(\sum_i |i\rangle - 2|a\rangle\right) \simeq |u\rangle + \frac{2}{\sqrt{N}}|a\rangle$$

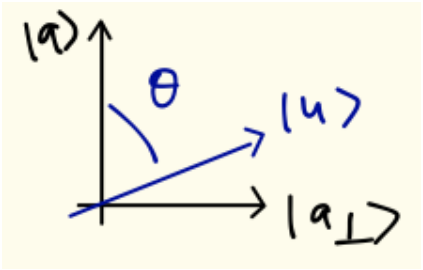
– closer to $|a\rangle$!

Claim: $\mathbf{R} \equiv \mathbf{W}\mathcal{O}$ is a *rotation* (in the N-dimensional $\mathcal{H}$) toward $|a\rangle$. (For now, we can focus on states with real coefficients, so we can draw a picture.) It is a rotation by 2α, with α defined above. Each of $\mathbf{W}$ and $\mathcal{O}$ is a reflection: $\mathbf{W}$ is a reflection about an unkown plane, and $\mathcal{O}$ reflects about a known plane. Because both $\mathbf{W}$ and $\mathcal{O}$ map vectors in the subspace spanned by $|a\rangle$ and $|u\rangle$ back into that subspace, we only need to worry about how $\mathbf{R}$ acts in the plane spanned by $|a\rangle$ and $|u\rangle$.

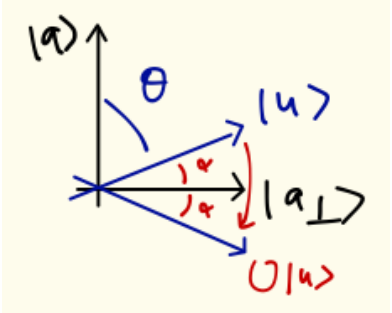
1. $|a\rangle, |u\rangle$ define a plane.



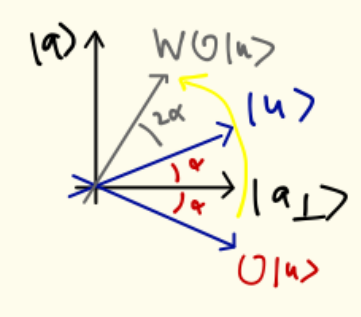
2. $|a_\perp\rangle$ is the vector orthogonal to $|a\rangle$ in this plane.



3. $\mathcal{O}$ acts as a reflection about $|a_\perp\rangle$ in this plane.



4. **W** acts as a reflection about $|u\rangle$ in the plane.



What happens if we act with $\mathbf{W}\mathcal{O}$ a second time? Note that the first step is still a reflection about $|a_\perp\rangle$, which rotates us the wrong way by 3α. But the second step is still a reflection about $|u\rangle$ (*not* about our vector $\mathbf{W}\mathcal{O}|u\rangle$); so this is a rotation by 5α in the right direction. In the end the second step (and every one that follows) is a rotation by 2α.

Therefore, if we do it many times, the overlap of the resulting state $(\mathbf{W}\mathcal{O})^k|u\rangle$ with $\langle a|$: $|\langle a|(\mathbf{W}\mathcal{O})^k|u\rangle|$ gets big as k gets big. (Big means close to 1.) More precisely:

$$g_k \equiv |\langle a|(\mathbf{W}\mathcal{O})^k|u\rangle| = \sin(2k+1)\alpha.$$

For small $g_0 = \sin\alpha = \frac{1}{\sqrt{N}} \sim \alpha$, the amplitude g_k reaches a maximum when $(2k+1)\alpha = \pi/2$, which means after $k \sim \frac{\pi}{4}\frac{1}{g_0} \sim \frac{1}{\sqrt{N}}$ iterations.

Notice that it doesn't work if $N = 2$. But if $N = 4$ it works perfectly on the first try! [Preskill 6.4.3] In that case $\sin\alpha = \frac{1}{\sqrt{N}} = \frac{1}{2}$ so $\alpha = 30°$. After one iteration, $\mathbf{W}\mathcal{O}|u\rangle$ makes an angle with $|a_\perp\rangle$ equal to $2\alpha + \alpha = 90°$ – it's the same as $|a\rangle$ for sure!

In fact for any N, there's a way to slick up the algorithm to make success a certainty.

Any kind of quantum computer that we could imagine making in the world will be build out of a set of elementary gates, like the ones described above, acting on a few qbits at a time.

Making the oracle and **W** from elementary gates is described in Preskill 6.4.1. The real importance of these quantum logic gates **H** and $\mathcal{C}_{\text{not}}$ that we've introduced is that *any* unitary operator acting on a collection of qbits can be constructed as a composition of these two kinds of elementary operations, on one or two qbits at a time. (On HW 6 you will show that you can make the uniform superposition $|u\rangle$ starting from a computational-basis state just using Hadamards.) We will not prove this result here; see the book by Schumacher (Chapter 18) for more on this.

The main obstacle to actually making a quantum computer is *decoherence* – it is a crucial assumption of the discussion above that the state of the computer does not become entangled

with its environment. As we will see in chapter 2, this is something that really wants to happen. There are lots of good ideas for getting around this, but it is a hard problem.

[End of Lecture 14]