14 Physics of Quantum Computing

Now is the time to take stock of what the previous chapters have shown. Simon's algorithm solved the first task that could be proved beyond polynomial-time classical algorithms, though it was relative to a hidden function. Shor's algorithm is concrete and solves factoring, but factoring hasn't been proved to lie outside deterministic polynomial time. Grover's algorithm can search among N items in $O(\sqrt{N})$ time, but where N is exponential in the number of qubits used to encode the search results. These are all advances over what is known or believed to be possible classically—and the quantum algorithms are available now.

Thus, in this chapter we take time to discuss how to achieve these advances physically. That is to say, what is involved in building a physical quantum computer? We will adopt the Dirac notation from physics to describe quantum states. We feel it important also to discuss interpretations of results that separate the quantum and classical worlds. Our own standpoint is that computational complexity holds a key not only to the separations but also to the interpretations. Complexity is central to the argument over *quantum supremacy*, which is perhaps better called by the mellower term *quantum advantage*:

> How can we tell when a quantum device has achieved a task that no classical device can feasibly emulate?

The goal of demonstrating quantum advantage has come to a boil in the six years since our first edition, with major companies, university centers, consortia, and smaller contenders all vying to be the first to achieve it definitively. We lay groundwork for understanding the physical side of this question and introduce notation and concepts that will serve the advanced algorithms in the rest of this text. We begin with a treatment of quantum reality that emphasizes how complex linear algebra is built around basic observable outcomes.

14.1 Coherence and Cards

If you wish to observe the suit of a face-down playing card, you can turn it over. The suit can be hearts ♡, diamonds ♢, clubs ♣, or spades ♠. Those are the basic outcomes. We distinguish them as **basis elements** by putting around them special notation invented by Paul Dirac: $|\heartsuit\rangle, |\diamondsuit\rangle, |\clubsuit\rangle, |\spadesuit\rangle$.

If a magician holds a standard playing card up to you so you cannot see the suit, you still have no doubt that it has one of those four suits. The magician

may trick you into thinking it has a different suit from when you saw it before, but it has a definite suit. It is a classical card. It does not change its suit or anything about its state. The deck is not magic.

The essence of quantum mechanics—for the finite discrete systems we consider—is that nature does provide special cards whose state is indeterminate. In this instance they behave as described by a vector $\boldsymbol{v}$ of four complex numbers (a, b, c, d) such that $|a|^2 + |b|^2 + |c|^2 + |d|^2 = 1$ according to some postulates, of which the first two are as follows:

(a) If you turn the card over, you get hearts with probability $|a|^2$, diamonds $|b|^2$, clubs $|c|^2$, and spades $|d|^2$.

(b) The magician can wave a wand over the card in the form of a 4×4 complex unitary matrix $\boldsymbol{M}$. The card then behaves as described by the vector $\boldsymbol{v}' = (a', b', c', d')$ such that $\boldsymbol{M}\boldsymbol{v} = \boldsymbol{v}'$.

These rules mean that the vector describes the **state** of the card, which can be written schematically as

$$|\kappa\rangle = a\,|\heartsuit\rangle + b\,|\diamondsuit\rangle + c\,|\clubsuit\rangle + d\,|\spadesuit\rangle\,.$$

The $|\cdot\rangle$ form, called a **ket**, signifies a column vector. We have used $\heartsuit, \diamondsuit, \clubsuit, \spadesuit$ rather than numbers inside the kets to emphasize that complex linear algebra can be built around all manner of basic observable outcomes.

These ideas extend to any number k of basis elements, not just $k = 4$. A **qubit** is defined by having $k = 2$, a **qutrit** by $k = 3$, and our cards are just called **quarts**. Whatever the arity, the probability rule (a) is similar. It is called the **Born rule** after Max Born. The behavior of (a) and (b) on the whole is called **coherence**, a term derived from physical waves. There is a third rule, which in our special case takes the following form:

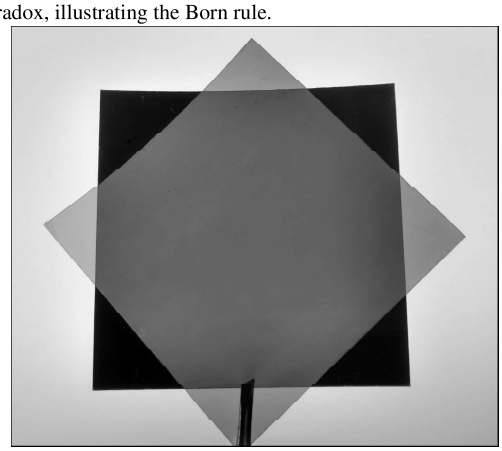
(c) If the observation in (a) gives hearts, then the card behaves as described by $(1, 0, 0, 0)$, which is the description of the basis vector $|\heartsuit\rangle$ itself. If diamonds, then the state becomes $(0, 1, 0, 0) = |\diamondsuit\rangle$; if clubs, then $(0, 0, 1, 0) = |\clubsuit\rangle$; if spades, then $(0, 0, 0, 1) = |\spadesuit\rangle$.

We will cover the general form of postulate (c) in section 14.5. *Why* quantum systems behave this way is called the **measurement problem**. We will duly touch on aspects that are commonly said to be counterintuitive and mysterious and "magic," but we will try to convince you that they are simply natural.

Indeed, to *see* quantum behavior all one needs are a few pieces of polarizing filter, such as used in sunglasses and cheaply available. Figure 14.1 shows two

pieces held with a tweezer in front of the LCD screen being used to write this chapter.

Figure 14.1
Three-polarizer paradox, illustrating the Born rule.



The diamond-oriented filter is between the horizontal filter and the screen. The corners of the horizontal filter are blocking all light from the screen. If the diamond filter were removed, the whole horizontal filter would block. The paradox is, how does putting another barrier between the horizontal filter and the light source *enable light to get through* in the center? To explain this, we use some new notation and a little trigonometry.

14.2 Dirac Notation

How do we get numbers out of the basis symbols? This requires first defining the **bra** form to go with **ket**. If $|\kappa\rangle = (a, b, c, d)$ as a column vector, then $\langle\kappa|$ is the row vector (a^*, b^*, c^*, d^*), where $*$ means complex conjugate. The point—note that bras and kets have points—is that making a "bra-ket" of basic outcomes gives 1 for like symbols and 0 otherwise. In our playing-card example,

$$\langle\heartsuit|\cdot|\heartsuit\rangle = \langle\diamondsuit|\cdot|\diamondsuit\rangle = \langle\clubsuit|\cdot|\clubsuit\rangle = \langle\spadesuit|\cdot|\spadesuit\rangle = 1,$$

$$\langle\heartsuit|\cdot|\diamondsuit\rangle = \langle\heartsuit|\cdot|\clubsuit\rangle = \langle\heartsuit|\cdot|\spadesuit\rangle = \langle\diamondsuit|\cdot|\clubsuit\rangle = \langle\diamondsuit|\cdot|\spadesuit\rangle = \langle\clubsuit|\cdot|\spadesuit\rangle = 0.$$

In general, a bra followed by a ket is an *inner product*. With $|\kappa\rangle$ as above and $|\lambda\rangle = e\,|\heartsuit\rangle + f\,|\diamondsuit\rangle + g\,|\clubsuit\rangle + h\,|\spadesuit\rangle$, we have

$$\langle\lambda| \cdot |\kappa\rangle = \langle\lambda, \kappa\rangle = e^*a + f^*b + g^*c + h^*d.$$

Having shown this, we will go back to writing inner products as $\langle\lambda, \kappa\rangle$.

A ket followed by a bra is an *outer product*. When we do this with basis elements we don't get 0 or 1; instead, we always get a new dimension marker. For example,

$$\begin{aligned} |\kappa\rangle\langle\lambda| \;=\; & ae^*\,|\heartsuit\rangle\langle\heartsuit| + af^*\,|\heartsuit\rangle\langle\diamondsuit| + ag^*\,|\heartsuit\rangle\langle\clubsuit| + ah^*\,|\heartsuit\rangle\langle\spadesuit| \\ +\; & be^*\,|\diamondsuit\rangle\langle\heartsuit| + bf^*\,|\diamondsuit\rangle\langle\diamondsuit| + bg^*\,|\diamondsuit\rangle\langle\clubsuit| + bh^*\,|\diamondsuit\rangle\langle\spadesuit| \\ +\; & ce^*\,|\clubsuit\rangle\langle\heartsuit| + cf^*\,|\clubsuit\rangle\langle\diamondsuit| + cg^*\,|\clubsuit\rangle\langle\clubsuit| + ch^*\,|\clubsuit\rangle\langle\spadesuit| \\ +\; & de^*\,|\spadesuit\rangle\langle\heartsuit| + df^*\,|\spadesuit\rangle\langle\diamondsuit| + dg^*\,|\spadesuit\rangle\langle\clubsuit| + dh^*\,|\spadesuit\rangle\langle\spadesuit|. \end{aligned}$$

It is usual to arrange the coefficients into a matrix:

$$|\kappa\rangle\langle\lambda| = \begin{bmatrix} ae^* & af^* & ag^* & ah^* \\ be^* & bf^* & bg^* & bh^* \\ ce^* & cf^* & cg^* & ch^* \\ de^* & df^* & dg^* & dh^* \end{bmatrix}.$$

When we take an outer product of a vector with itself, there is a special significance when we multiply the resulting matrix by another ket-vector and use associativity:

$$(|\kappa\rangle\langle\kappa|)\,|\lambda\rangle = |\kappa\rangle\,((\langle\kappa|)\,|\lambda\rangle) = |\kappa\rangle\langle\kappa, \lambda\rangle = c\,|\kappa\rangle,$$

where c is the scalar value $\langle\kappa, \lambda\rangle$. Thus, $|\kappa\rangle\langle\kappa|$ is an *operator* that when applied to another vector gives the projection of that vector onto the span of $|\kappa\rangle$. It is generally *not* unitary: the projection is not one to one. We say more about projections when defining *measurements* formally in section 14.5.

The bra-ket notation widens to accommodate operators. The abstract form

$$\langle\kappa|\ \boldsymbol{M}\ |\lambda\rangle$$

is sometimes called the *triple product* but is really just ordinary matrix-vector multiplications to get a scalar. If we identify an n-qubit quantum circuit C with the unitary operation it computes, and if we consider basic inputs $x, z \in \{0,1\}^n$, then

$$\langle z|\,C\,|x\rangle$$

gives the amplitude of the event of C on input x giving output z. In terms of the length-2^n complex state vector obtained after giving x as input to C, it is the complex value in position z. Note that z is on the left since we apply operators to column vectors on the right, though this is opposite to our tendency to picture circuits as running from left to right.

Complex conjugation flips ket and bra and reverses the right-to-left sequence order. If C breaks into unitary operators as $C = U_m U_{m-1} \cdots U_1$, then

$$\begin{aligned} \langle z| C |x\rangle^* &= \langle z| U_m U_{m-1} \cdots U_1 |x\rangle^* = \langle x| (U_m U_{m-1} \cdots U_1)^* |z\rangle \\ &= \langle x| U_1^* \cdots U_{m-1}^* U_m^* |z\rangle . \end{aligned}$$

We complete our treatment of Dirac notation by noting that $|\kappa\rangle |\lambda\rangle$ standardly denotes the tensor product $|\kappa\rangle \otimes |\lambda\rangle$. We defined tensor products in chapter 3, so we simply remark here that, in our running example, $|\kappa\rangle |\lambda\rangle$ is a vector of length 16 rather than a 4×4 matrix. In presenting Shor's algorithm and before, we wrote a functional superposition state as a function:

$$\boldsymbol{a}[xy] = \begin{cases} \frac{1}{\sqrt{2^n}} & \text{if } f(x) = y; \\ 0 & \text{otherwise.} \end{cases}$$

Now we write it as

$$\frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle |f(x)\rangle .$$

The meaning is the same; the choice depends on how it will be used.

14.3 What Are Qubits?

We have talked about qubits all through this text. In chapter 3 we derived them from linear algebra and Euclidean distance. Above in section 14.1 we gave postulates for their behavior—now this means having $k = 2$ with numbers a, b such that $|a|^2 + |b|^2 = 1$. But what *is* a qubit? Nature provides multiple entities that behave as qubits. We emphasize qubit*s*, plural, because one needs a source of multiple qubits in the same initial state to verify that their behavior conforms to the stated probabilities of the Born rule.

What was the first physical device to provide proof? Historically it was a magnetic device built by Otto Stern and Walter Gerlach in 1922. It beamed silver atoms through a magnetic field that, despite the atoms not being electrically charged, deflected them according to their *spin*—that is, each atom's

angular momentum. If the angular momentum values were classical values on a continuous scale, then the pattern of atoms landing on a detector would be continuous. Instead, there are just two landing places, one of them "down" ($\downarrow$) the other "up" ($\uparrow$), as shown in figure 14.2.

The finest point is that, *as far as we know*, the atoms are identical as they enter the magnetic field. Some go up, and some go down, in ways we cannot tell in advance. We can attenuate the beam so that most of the time only a single atom is traveling through the device. We still cannot tell as it enters which path it will later take. There may be some unknown factors that determine the path, but—and this is also important—our methods of preparing the atoms for their flight are not knowingly biased in regard to them. We say more about possible "hidden variables" in sections 14.7 and 14.10. The point for now is not what the atoms "are" but what they *do*:

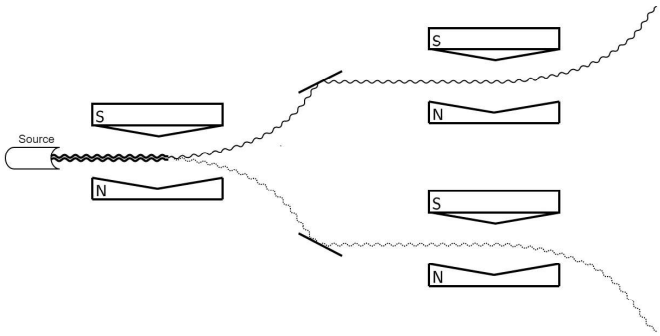
- We have a device that allows us repeatedly to sample results by which we get the outcome $\uparrow$ with some probability p and $\downarrow$ with probability $1-p$.
- We can prepare the atoms to give a known value p, for instance, $p = 0.5$, from a configurable set of possible values. We can observe that, when we shoot many atoms, almost always the frequency of those giving the outcome $\uparrow$ is close to p.
- The atoms behave as if controlled by complex numbers a, b such that $|a|^2 = p$ and $|b|^2 = 1-p$. We can also prepare those numbers, but we cannot examine a flying atom and tell what a and b it has unless we already know. We can write the state in Dirac notation as

$$a\,|\uparrow\rangle + b\,|\downarrow\rangle\,.$$

- A reason we know that the (a, b) mechanism is operational is that we can apply operations that are representable as multiplying (a, b) by a 2×2 complex *unitary* matrix M to get (a', b'), so that the resulting atoms give $\uparrow$ with probability $p' = |a'|^2$ and $\downarrow$ with probability $1 - |b'|^2$.
- A reason we know that the (a, b) notion of "state" is intrinsic is that, if we take the beam of atoms that go down and feed it into a second Stern-Gerlach device, we don't see a split. The beam stays down: they behave as described by the state $(0, 1)$, which always gives $\downarrow$. This is also shown in figure 14.2.

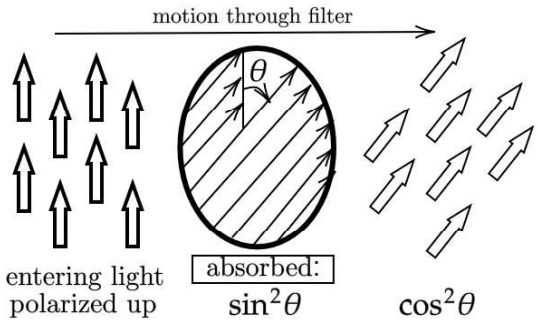
Happily, one does not need to visit a Stern-Gerlach device to see the Born rule in action. The polarizing filters shown in figure 14.1 suffice, as does using a third filter or sunglasses in place of the computer screen. Light emerging

Figure 14.2
Cascaded Stern-Gerlach devices.



from a filter has waves that oscillate in only one direction in a plane seen head-on. Let's set the first filter so that direction is vertical, as it is for the cscreen in figure 14.1. Now set the next filter at an angle θ from the vertical, as shown in figure 14.3.

Figure 14.3
Born rule for polarized light.



Let us use the Dirac notation $|\uparrow\rangle$ for the vertically up polarized state. The second filter imposes its own basis on the light it allows to pass through. We can denote its basis by $|\nearrow\rangle$ for the direction at angle θ and $|\nwarrow\rangle$ for the axis orthogonal to θ (oriented 90 degrees counterclockwise looking through the filter). When we express the incoming light's vector $|\uparrow\rangle$ using the second filter's basis coordinates, we get:

$$|\uparrow\rangle = \cos(\theta)\,|\nearrow\rangle + \sin(\theta)\,|\nwarrow\rangle\,.$$

The second filter effects a **measurement** with the following physical rules. If the outcome is $|\nearrow\rangle$ then the photon passed through. If it is $|\nwarrow\rangle$, however, the photon was *absorbed*. The Born rule gives us the probabilities:

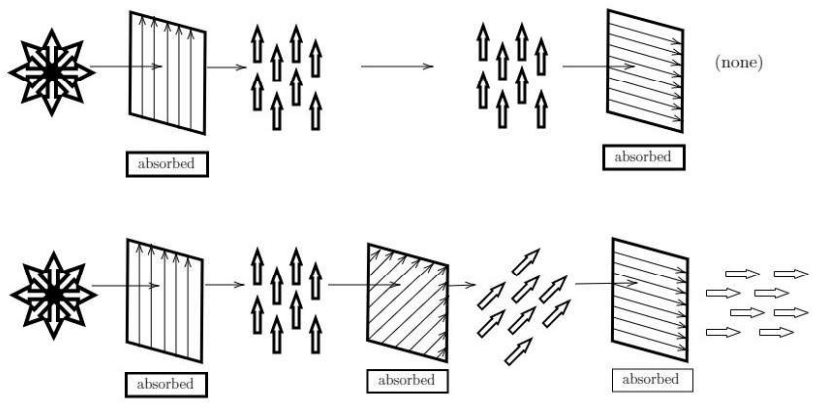
$$\begin{aligned} \Pr(|\nearrow\rangle), \text{ i.e., } \Pr[\textit{pass through}] &= \cos^2(\theta); \\ \Pr(|\nwarrow\rangle), \text{ i.e., } \Pr[\textit{blocked}] &= \sin^2(\theta). \end{aligned}$$

The only thing we lose compared to the Stern-Gerlach device is the ability to see the phenomenon with individual particles. With a massive number of photons, however, we can see the probabilities play out with our eyes: the portion of light passing through is $\cos^2(\theta)$. Varying θ by twirling the filter in front of the LCD screen shows how smoothly the trigonometric law operates. And when $\theta = 90^\circ$, giving $\cos(\theta) = 0$, the blocking is quite close to total, as shown by the black corners shown in figure 14.1. Likewise, sunglasses use vertical polarization to filter out glare from shiny surfaces, which is mostly horizontally polarized.

Now we can explain the paradox from figure 14.1. Figure 14.4 shows it schematically. The light arriving at the first filter is unpolarized, meaning it is an equal mixture of all polarizations. Relative to figure 14.1, the first filter is the LCD screen and the middle filter is the diamond one at angle $\theta = 45^\circ$. By the Born rule, $\cos^2(45^\circ) = \frac{1}{2}$ of the light gets through the diamond. This is evident by looking at the corners of the diamond compared to the screen. The horizontal filter is then at an angle θ' relative to the middle filter.

Figure 14.4

The three-polarizer paradox. Shading indicates portions absorbed or passed through.



Here we again have $\theta' = 45^\circ$ and again half of the incoming light passes through. The upshot is that $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ of the light from the screen comes through. Again this is evident from figure 14.1.

If we twirl the middle filter while holding steady the third filter (i.e., the horizontal filter in the foreground), then we obtain varying angles θ and $\theta' = 90^\circ - \theta$. The portion of light coming through is

$$\cos^2(\theta)\cos^2(\theta') = \cos^2(\theta)\sin^2(\theta).$$

This is maximum when $\theta = \theta' = 45^\circ$ and tails off to 0 both as $\theta \to 0^\circ$ and $\theta \to 90^\circ$. In fact, this tells exactly how an LCD display works in the first place. For a screen that emits vertically polarized light as shown, the light source at the back is *horizontally* polarized, while the screen's surface is set vertical. Playing the middle-filter role is a layer of tiny liquid crystals that can be rotated by electric charges. Each pixel uses three crystals, one stationed in front of a red element (R), the second green (G), and the third blue (B). Different triads $\theta_1, \theta_2, \theta_3$ of rotation angle for each pixel give different RGB combinations for the pixel color that our eyes perceive. Zero charge gives $\theta_1 = \theta_2 = \theta_3 = 0$ and blocks all light, which is why 000000 is the hexadecimal code for black. Other LCD displays may use transverse 45 degree angles for source and screen. This allows a person wearing sunglasses to use the screen in either portrait or landscape mode.

Our speaking of $|\nearrow\rangle, |\nwarrow\rangle$ as a "separate basis" may have seemed a fussy detail. We could have spoken in terms of θ directly without that bit of bookkeeping. However, the great issue between Shor's algorithm and whether the factoring problem is in classical polynomial time can be framed as whether our classical notation can keep up with nature's inherent efficiency. Thus, the efficacy of our bookkeeping is an object of analysis.

In linear algebra we can freely transform vectors to other bases. What our polarized-light example already conveys is that the transformed basis can become another physically viable measurement target. For example,

$$|+\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) \qquad \text{and} \qquad |-\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle)$$

form another frequently used target basis. This is the **Bell basis**, which was mentioned first in section 8.3 of chapter 8 and is employed further below beginning in section 14.5.

Staying with our original basis, if we have a second particle with state $c\,|\uparrow\rangle + d\,|\downarrow\rangle$, then we can regard our two particles side by side as

$$(a\,|\uparrow\rangle + b\,|\downarrow\rangle) \bullet (c\,|\uparrow\rangle + d\,|\downarrow\rangle).$$

Here the $\bullet$ is a loud use of our usual computer science notion of concatenation. The corresponding tensor product in Dirac notation is

$$(a\,|\uparrow\rangle + b\,|\downarrow\rangle) \otimes (c\,|\uparrow\rangle + d\,|\downarrow\rangle) = ac\,|\uparrow\uparrow\rangle + ad\,|\uparrow\downarrow\rangle + bc\,|\downarrow\uparrow\rangle + bd\,|\downarrow\downarrow\rangle\,.$$

Like concatenation, tensor product is not commutative, so it is as if we are singling out one of the particles as going first. Now we have four basis states $|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle$. The vector (ac, ad, bc, bd) over this basis is separable as defined in chapter 3, that is, decomposable as a tensor product of shorter vectors. It is still a unit vector: you can check that $|ac|^2 + |ad|^2 + |bc|^2 + |bd|^2 = 1$.

The final property is that we can freely use unit vectors in the larger space, as follows.

DEFINITION 14.1 A **quantum register** of k qubits $q_1, \ldots, q_k$ is represented by complex unit vectors

$$(a_0, \ldots, a_{K-1})$$

with $K = 2^k$ that obey the k-qubit Born rule that, for each $J < K$ in binary notation as $J = b_1 b_2 \ldots b_k$,

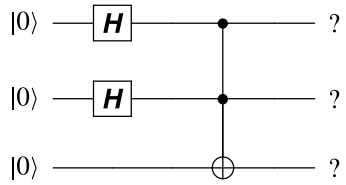
$$|a_J|^2 = \Pr[q_1 = |b_1\rangle \wedge q_2 = |b_2\rangle \wedge \cdots \wedge q_k = |b_k\rangle].$$

Dirac notation allows us to write the whole outcome as

$$|J\rangle = |b_1 b_2 \cdots b_k\rangle = |b_1\rangle\,|b_2\rangle \cdots |b_k\rangle = |b_1\rangle \otimes |b_2\rangle \otimes \cdots \otimes |b_k\rangle\,.$$

This is more structured than having a single K-level entity with basis states $|0\rangle$ through $|K-1\rangle$ because of the tensor products. Now the basis states give properties of binary strings in $\{0,1\}^k$ and can convey entanglement. Nevertheless, the intent is that a quantum register can be treated as a unit. Operations and entanglements *between* registers can be singled out for emphasis.

In chapter 7 we showed how to entangle two qubits using one Hadamard gate and one ***CNOT*** gate. Here we exemplify an entangled quantum system with $k = 3$ using two Hadamards and a Toffoli gate:



We started with three separate qubits each in the basis state $|0\rangle = (1,0)$. So the whole initial state is $|000\rangle$. We first applied the Hadamard gates to the first and second qubits to put them in the state

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

We still have three separated qubits, collectively in the state $|+\rangle \otimes |+\rangle \otimes |0\rangle$, which can be written $|{+}{+}0\rangle$. Then we apply the Toffoli gate, and what we get is

$$\frac{1}{2}(|000\rangle + |100\rangle + |010\rangle + |111\rangle).$$

Taking care to compose $\boldsymbol{H} \otimes \boldsymbol{H}$ and ***TOF*** right to left, the entire eight-dimensional computation is

$$\begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0 \end{bmatrix} \cdot \frac{1}{2} \begin{bmatrix} 1&0&1&0&1&0&1&0\\ 0&1&0&1&0&1&0&1\\ 1&0&1&0&1&0&1&0\\ 0&1&0&1&0&1&0&1\\ 1&0&1&0&-1&0&-1&0\\ 0&1&0&1&0&-1&0&-1\\ 1&0&1&0&-1&0&1&0\\ 0&1&0&1&0&-1&0&1 \end{bmatrix} \cdot \begin{bmatrix} 1\\0\\0\\0\\0\\0\\0\\0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1\\0\\1\\0\\1\\0\\0\\1 \end{bmatrix}.$$

The third qubit depends on the conjunction of the first two. It will give outcome $|1\rangle$ if each of the first two gives outcome $|1\rangle$ in its respective place, but otherwise the third will give outcome $|0\rangle$. The outcomes $|000\rangle$, $|010\rangle$, $|100\rangle$, and $|111\rangle$ each have probability $1/4$, whereas the other four outcomes cannot happen.

The next property is that we don't have to measure all the qubits but can measure just some of them or even measure in a different way—that is, in a different basis. This needs its own section for definitions, but let's continue our example here. Suppose we measure the third qubit and get the outcome $|0\rangle$. This outcome had probability $3/4$, but what can we say about the state of the

system now? The answer is that its state becomes

$$\frac{1}{\sqrt{3}}(|000\rangle + |100\rangle + |010\rangle) = \frac{1}{\sqrt{3}}(|00\rangle + |10\rangle + |01\rangle) \otimes |0\rangle\,.$$

It is, moreover, still entangled: the state $\frac{1}{\sqrt{3}}(|00\rangle + |10\rangle + |01\rangle)$ cannot be written as a tensor product of single-qubit states.

Where did the 3 come from? The Dirac notation helps us track the possibilities more compactly than an 8-vector with three entries $1/\sqrt{3} = 0.57735\ldots$ might, but it leaves even more puzzlement on *why* than the two-qubit entanglement in chapter 7.

14.4 Transformations and the Bloch Sphere

In the last section we represented a qubit as

$$a\,|0\rangle + b\,|1\rangle\,,$$

where a and b are complex numbers. Although we referred to (a,b) as *the* state, there is an equivalence relation under which nature makes no distinctions. Define (a,b) and (a',b') to be equivalent if there is a unit complex number c such that $a' = ca$ and $b' = cb$. Then $|a'|^2 = |a|^2$ and $|b'|^2 = |b|^2$, so the probabilities are the same, but the difference in complex phase is the same between a' and b' as between a and b. So if we take (a,b) as representing this equivalence class, we may as well assume that the phase of a is zero, which means that a is a nonnegative real number. Let φ be the phase angle of b.

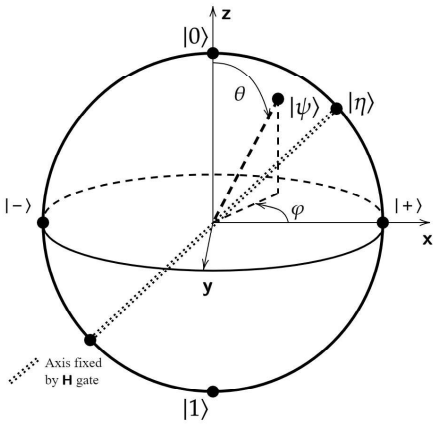
This leaves the two real numbers b_1, b_2 used to write $b = b_1 + b_2 i$. They are constrained by the requirement that $1 = |a|^2 + |b|^2 = a^2 + b_1^2 + b_2^2$. So we can represent the entire quantum state by the following:

- A real number between 0 and 1, without loss of generality of the form $a = \cos(\theta/2)$, where $0 \le \theta \le \pi$
- The phase angle φ, which makes $b = e^{i\varphi}\sin(\theta/2)$

The angle θ is a latitude, and the angle φ is a longitude. Thus, we have mapped states in one-to-one fashion to points on the surface of a sphere. This is called the **Bloch sphere** after Felix Bloch. The latitude is reckoned down from the north pole, so the north pole is $\theta = 0$ and the south pole is $\theta = \pi$. Thus, we have the following:

- The north pole is $\theta = 0$, so $a = \cos(0) = 1$ and $\sin(0) = 0$ so $b = 0$, making φ immaterial. So the north pole is $|0\rangle$.
- The south pole has $\theta = \pi$, so $a = \cos(\pi/2) = 0$ and $b = e^{i\varphi}$. The longitude is immaterial at the pole, so this is $|1\rangle$.
- The equator has $a = \frac{1}{\sqrt{2}}$ and $b = e^{i\varphi}\frac{1}{\sqrt{2}}$. If we view $\varphi = 0$ at far right, then the state $|+\rangle$ is there and the state $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ is at the far left.

Figure 14.5
The Bloch sphere with axes and corresponding operators.



It is important to note the difference from our usual way of representing the orthogonal vectors $(1,0)$ and $(0,1)$ in the (real number) plane. Those are at 90 degrees from each other, whereas $|0\rangle$ and $|1\rangle$ on the Bloch sphere are 180 degrees apart. Still, we can put rectangular coordinates on the sphere. The axis going from $|-\rangle$ on the left to $|+\rangle$ on the right can still be called the x-axis. Let us call the other equatorial axis, going from $\frac{1}{\sqrt{2}}(1,i)$ in front to $\frac{1}{\sqrt{2}}(1,-i)$ in back, the y-axis, and the vertical one through the poles the z-axis. This is shown in figure 14.5.

The ***NOT*** gate flips the poles. It also sends $\frac{1}{\sqrt{2}}(1,i)$ to $\frac{1}{\sqrt{2}}(i,1)$, but remember that up to equivalence the latter is the same as $\frac{1}{\sqrt{2}}(1,-i)$, so we have interchanged our equatorial front and back. Thus, the effect on the Bloch sphere is a 180-degree rotation around the x-axis. This is why the gate is called σ_x, or simply ***X***.

The gate $\boldsymbol{Y}$ also flips the poles and this time swaps $|+\rangle$ and $|-\rangle$ while fixing our front and back points. This means it rotates 180 degrees around the y-axis. The gate $\boldsymbol{Z}$ fixes the poles and rotates 180 degrees around the z-axis, turning the equator. The gate $\boldsymbol{S}$ is an equatorial rotation by 90 degrees (note $\boldsymbol{S}^2 = \boldsymbol{Z}$) and $\boldsymbol{T}$ by 45 degrees.

Now consider the point $\theta = \frac{\pi}{4}$, $\varphi = 0$, which is at 45 degrees north on the prime meridian. The corresponding state vector is $|\eta\rangle = \left(\cos\left(\frac{\pi}{8}\right), \sin\left(\frac{\pi}{8}\right)\right)$. It takes some labor with trigonometric identities to see that this is fixed by the Hadamard matrix $\boldsymbol{H}$. Since $\boldsymbol{H}$ sends the north pole to $|+\rangle$, this means that $\boldsymbol{H}$ rotates the Bloch sphere 180 degrees around the diagonal axis through the center and $|\eta\rangle$. This is also shown in figure 14.5. In fact, every 2×2 unitary matrix gives a rotation on the Bloch sphere.

Another fact about the Bloch sphere is that *latitude gives the probability*. Here we count latitude as 1 at the north pole, 0.5 at the equator, and 0 at the south pole. For angle θ away from the north pole, the latitude is $\frac{1+\cos(\theta)}{2}$. By a standard trigonometric identity, this is $\cos^2(\theta/2)$, which means it is a^2, which means it is the probability of getting $|0\rangle$.

Connected to this is that points *inside* the Bloch sphere have statistical meaning. The center of the sphere equals $0.5\,|0\rangle + 0.5\,|1\rangle$ and represents the uniform *classical* probability distribution over the two basic outcomes. A general distribution $p\,|0\rangle + (1-p)\,|1\rangle$ has latitude p by the above reckoning. Hence we reach this interpretation:

- The points on the z-axis through the poles give all the purely classical distributions on $\{|0\rangle, |1\rangle\}$.
- The points on the surface of the sphere give all the "purely quantum" distributions on $\{|0\rangle, |1\rangle\}$.
- All other points inside the sphere have the form $\boldsymbol{\rho} = p(\theta, \varphi) + (1-p)(\pi - \theta, \pi + \varphi)$ and hence represent classical distributions between a pure state $|\psi\rangle$ and its antipode on the surface, which is the pure state orthogonal to $|\psi\rangle$.

All classical distributions over points on the surface—except the distribution giving probability 1 to one point—yield a point inside the sphere. Those are the *mixed states* of one qubit. The sphere plus the points inside are called the **Bloch ball**. Points on the surface are typified by the spin states discussed above, while we discuss below the physical meaning of points inside.

The idea behind the Bloch sphere extends to qutrits, quarts, indeed d-ary **qudits** for any d, and to systems of more than one qubit or qudit. The geometry is not as neat as a sphere, however, except that the analogue of the *Bloch ball* is always a convex body. This enables the definition of mixed state to be the same for all systems, as follows:

DEFINITION 14.2 A **mixed state** is a convex combination of pure states.

In section 14.6 we give a canonical representation for mixed states, after first discussing measurements of pure states in section 14.5. Before moving on, however, we need to discuss some points about notation.

We have promoted the idea that the Dirac notation takes a targeted outcome or attribute u and denotes a corresponding quantum state vector by wrapping a ket around it to make $|u\rangle$. Previously we denoted vectors in boldface, such as $\boldsymbol{u}$. In this section, we have taken Greek letters ϕ, ψ, η that already stood for quantum state vectors and wrapped kets around them as $|\phi\rangle, |\psi\rangle, |\eta\rangle$ to make them "look more quantum." Usually this redundancy—which is often seen in the literature—is harmless. It needs comment to avoid problems in these contexts, however:

- *Duals*. Our previous notation denotes the conjugate transpose of a vector $\boldsymbol{u}$ by $\boldsymbol{u}^*$. In Dirac notation, it becomes $\langle u|$ without writing a star. Writing $\langle u^*|$ would be like a double negative. One has to remember to conjugate the contents when expanding expressions involving $\langle u|$.
- *Outer products*. We have not needed to consider outer products previously in this textbook. Thus, we were able to duck the issue that there seems to be no standard non-Dirac notation for them. Some authorities write $\boldsymbol{u} \otimes \boldsymbol{v}$, but that would cause confusion with tensor products. We have already shown in section 14.2 how handy Dirac notation is not only for defining outer products $|u\rangle\langle v|$ but also for combining them with inner products and matrices. This sets up the temptation to write $|\boldsymbol{u}\rangle\langle\boldsymbol{v}|$ for outer product when $\boldsymbol{u}$ and $\boldsymbol{v}$ already stand for vectors.
- *Vectors as targets*. If the objects we seek already are vectors, such as when trying to solve $\boldsymbol{u} = \boldsymbol{A}^{-1}\boldsymbol{x}$ in chapter 18, then $|\boldsymbol{u}\rangle$ means "a quantum state that denotes $\boldsymbol{u}$." It may not be $\boldsymbol{u}$ itself but a vector in a larger Hilbert space.

We will attend to these issues as they come. They are a reason we avoided Dirac notation early on. The outer-product convenience leads us to excuse the

redundancy of writing $|\boldsymbol{u}\rangle$ when it is the same as $\boldsymbol{u}$. Even in the third case, $|\boldsymbol{u}\rangle$ will project to $\boldsymbol{u}$—or to some approximation of $\boldsymbol{u}$—in a well-defined sense.

With *mixed states*, however, there is a clear taxonomy: some arise as single outer products, and others do not. Outer products are matrices that denote linear operators. Accordingly, we resume our bold notation for matrices and operators, writing $\boldsymbol{\rho}$ for a typical mixed state. We will enunciate early in section 14.6 the possibility that outer products and not vectors should be regarded as *the* basic quantum elements.

14.5 Measurements of Pure States

The 2×2 Hadamard matrix $\boldsymbol{H}$ also effects the change from the standard $|0\rangle, |1\rangle$ basis to the basis $|+\rangle, |-\rangle$. This works in reverse because $\boldsymbol{H}^2 = \boldsymbol{I}$. This also enables changing our notion of how we measure away from fixing on the standard basis as this text has done to here. We can define measurement as an *operator*, conveniently using the Dirac notation.

To see the idea, consider the standard measurement of $|\kappa\rangle = a|0\rangle + b|1\rangle$. The probability p_0 of getting $|0\rangle$ falls out of a triple product of $\langle\kappa|$ and $|\kappa\rangle$ with the operator defined as the outer product of $|0\rangle$ with itself—really with $\langle 0|$, which is its dual:

$$\langle\kappa| \cdot |0\rangle \langle 0| \cdot |\kappa\rangle = (\langle\kappa, 0\rangle)(\langle 0, \kappa\rangle) = (a^*)(a) = |a|^2$$

The state after getting that result is of course just $|0\rangle$, but we can say more generally how it happens:

$$\frac{1}{\sqrt{p_0}} |0\rangle \langle 0| |\kappa\rangle = \frac{1}{a} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = |0\rangle$$

Now suppose we want to measure $|\kappa\rangle$ in the $|+\rangle, |-\rangle$ basis. For the outcome $|+\rangle$ we can imitate the above form:

$$\langle\kappa| \cdot |+\rangle \langle +| \cdot |\kappa\rangle = (\langle\kappa, +\rangle)(\langle +, \kappa\rangle) = (a^*)(a) = |a|^2$$

Notice that because $|+\rangle = \boldsymbol{H}|0\rangle = \langle 0|\boldsymbol{H}$, this is the same computation as

$$\langle\kappa|\boldsymbol{H}|0\rangle \langle 0|\boldsymbol{H}|\kappa\rangle = \langle\kappa|\boldsymbol{H} \cdot |0\rangle \langle 0| \cdot \boldsymbol{H}|\kappa\rangle .$$

The latter gives the same result as measuring $\boldsymbol{H}|\kappa\rangle$ in the $|0\rangle, |1\rangle$ basis to get the probability for the outcome $|0\rangle$. It helps here that $\boldsymbol{H}$ is *Hermitian*, that is,

$\boldsymbol{H} = \boldsymbol{H}^*$. The point of the former is that $|+\rangle\langle+|$ creates an operator. Now we involve the other outcomes and observe that

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

The property of the outer products summing to the identity is preserved under any change of basis transformation. This fact undergirds the following abstraction. A matrix $\boldsymbol{P}$ is **positive semidefinite** (PSD) if there is a matrix $\boldsymbol{A}$ such that $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{A}^*$; it is a **projection** if also $\boldsymbol{P}^2 = \boldsymbol{P}$. Matrices of the form $|\phi\rangle\langle\phi|$ are projections, as we showed in section 14.2.

DEFINITION 14.3 A **projective measurement** is given by a set $\{\boldsymbol{P}_1, \ldots, \boldsymbol{P}_m\}$ of projections such that

$$\sum_{j=1}^{m} \boldsymbol{P}_j = \boldsymbol{I}.$$

After measuring a (pure) state $|\phi\rangle$ the system selects some j with probability $p_j = \langle\phi|\ P_j\ |\phi\rangle$ and transits to the state $\frac{1}{\sqrt{p_j}} P_j\,|\phi\rangle$.

For a schematic example, let us revisit our "quart" system from section 14.1 with $|\heartsuit\rangle, |\diamondsuit\rangle$ for the red playing card suits and $|\clubsuit\rangle, |\spadesuit\rangle$ for the black suits. Suppose we wish to measure for the outcomes "red," that is, $|\heartsuit\rangle$ or $|\diamondsuit\rangle$, versus "black." The corresponding projectors are

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The probability of "red" given $|\kappa\rangle = a\,|\heartsuit\rangle + b\,|\diamondsuit\rangle + c\,|\clubsuit\rangle + d\,|\spadesuit\rangle$ is

$$\langle\kappa|\ P_1\ |\kappa\rangle = \langle(a^*, b^*, 0, 0), (a, b, c, d)\rangle = |a|^2 + |b|^2.$$

The probability of "black" is similarly $|c|^2 + |d|^2$. If the outcome is "red," the next state is the former of these two, else the latter:

$$|\text{red}\rangle = \frac{a\,|\heartsuit\rangle + b\,|\diamondsuit\rangle}{\sqrt{|a|^2 + |b|^2}} \qquad \text{or} \qquad |\text{black}\rangle = \frac{c\,|\clubsuit\rangle + d\,|\spadesuit\rangle}{\sqrt{|c|^2 + |d|^2}}$$

Alternatively, we could define a measurement with the same P_1 but with P_2 singling out clubs and P_3 spades. If the outcome is red then the state is the same as before, but otherwise the state would become definitely $|\clubsuit\rangle$ or definitely $|\spadesuit\rangle$. The difference from the state $|\text{black}\rangle$ might seem arbitrary, with $|\text{black}\rangle$ signifying only ignorance of the "true property" $|\clubsuit\rangle$ or $|\spadesuit\rangle$, but the difference is implied by how we defined the measurement. Whether we could *engineer* the measurement to preserve coherence when giving the outcomes red or black is another matter. In the next section we show what preserving coherence involves. We could instead try to engineer a measurement whose outcomes are "major suit" (meaning $\heartsuit$ or $\spadesuit$) versus "minor suit."

Now consider the corresponding general pure state of a two-qubit system: $|\phi\rangle = a\,|00\rangle + b\,|01\rangle + c\,|10\rangle + d\,|11\rangle$. There is extra structure here from *non*-entangled states breaking into tensor products, such as the basis state $|00\rangle$ being $|0\rangle \otimes |0\rangle$ and the case $a = b = c = d = \frac{1}{2}$ being $|+\rangle \otimes |+\rangle$. The outcome "red" now equates to "the first qubit is zero" and corresponds to $|0\rangle \otimes |+\rangle$. The outcome "major suit" equates to "$|00\rangle$ or $|11\rangle$," which doesn't isolate a value for either qubit, and the corresponding two-qubit state is entangled. It does, however, isolate a value for the first pair of entries in the **Bell basis** of *two* qubits versus the second pair:

$$\begin{aligned}
|\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\
|\Phi^-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\
|\Psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \\
|\Psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)
\end{aligned}$$

Thus, when "major suit" is transformed to this basis, the structure is the same as with "red" versus "black," which corresponded to measuring the first qubit in the standard basis.

Whatever one's interpretation of the mechanism for producing the next state after a projective measurement, the physical fact of the projection is undeniable. We can revisit the polarizing filter example. Each filter effects on each of myriad passing photons a projective measurement whose outcomes can be called "success: pass through" or "failure: absorbed." The success case projects the photon onto the filter's axis. When the angle θ of the next filter is 90 degrees, the probability of success is 0. But when the third filter is inserted

at $\theta' = 45$ degrees, half of the photons that passed through the first filter succeed and are then projected along θ'. They then encounter the last filter at θ which is 45 degrees from θ', and again there is a $1/2$ probability of success. The actual wave dynamics are more complicated, but the results are accurately described by this simple figuring of repeated measurements.

14.6 Mixed States and Decoherence

We reintroduce our old friends Alice and Bob. Well, for now just Alice—Bob will be way off in the distance. She buys a qubit from a vendor who says it is in the pure $|+\rangle$ state, having been prepared from $|0\rangle$ by applying $\boldsymbol{H}$. To back up the claim, the vendor measures a slew of other similarly prepared qubits in the $|0\rangle, |1\rangle$ basis, and Alice sees that the outcomes have close to a 50-50 split.

Alice wishes to measure it in the $|+\rangle, |-\rangle$ basis. Actually, what she intends is to apply $\boldsymbol{H}$ to it and measure in the standard basis. She expects to receive the result $|0\rangle$. There are two ways she can be disappointed. Our point will be how they are alike.

First, the qubit might instead be in the mixed state of a 50-50 classical split between $|0\rangle$ and $|1\rangle$. This would produce the same statistics from the vendor. We can calculate expressly what will happen when Alice applies $\boldsymbol{H}$ to it and measures. To motivate the representation of mixed states, we first note a consequence of the last section about measuring pure states that follows from the identity

$$\langle\kappa| \left(|\phi\rangle \langle\phi|\right) |\kappa\rangle = \langle\phi| \left(|\kappa\rangle \langle\kappa|\right) |\phi\rangle ,$$

which in turn follows from associativity and the scalar multiplication of $\langle\kappa, \phi\rangle$ and its dual $\langle\phi, \kappa\rangle$ being commutative:

> Every measurement of a pure state $|\kappa\rangle$ involves only the outer product $|\kappa\rangle \langle\kappa|$. Hence, $|\kappa\rangle \langle\kappa|$ embodies all knowledge we can gain about the state.

The consequence is that the density matrix, defined as follows, gives all quantum information about pure states as well as mixed states.

DEFINITION 14.4 The **density matrix** of a mixed state given as a convex combination of pure states $|\phi_k\rangle$ with nonnegative real coefficients p_k summing to 1 is

$$\rho = \sum_k p_k |\phi_k\rangle \langle\phi_k| .$$

A density matrix ρ designates a pure state if and only if $\rho^2 = \rho$. A second important fact about density matrices ρ is that they are always Hermitian: $\rho^* = \rho$. This is because

$$(|\phi_k\rangle \langle\phi_k|)^* = (\langle\phi_k|^*)(|\phi_k\rangle)^* = |\phi_k\rangle \langle\phi_k|$$

for each k. Hence, they can be decomposed in a canonical way according to the following theorem.

THEOREM 14.5 For every $N \times N$ Hermitian matrix $\boldsymbol{A}$, we can find an orthonormal basis of eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$ with associated real eigenvalues $\lambda_1, \ldots, \lambda_N$ such that

$$\boldsymbol{A} = \lambda_1 |\boldsymbol{u}_1\rangle \langle\boldsymbol{u_1}| + \cdots + \lambda_N |\boldsymbol{u}_N\rangle \langle\boldsymbol{u_n}|. \tag{14.1}$$

If each λ_j is unique, then so are the $\boldsymbol{u}_j$. If some eigenvalue is duplicated, for example, $\lambda_{j_1} = \cdots = \lambda_{j_r}$, then all ways of picking an orthonormal basis $\{\boldsymbol{u}_{j_1}, \ldots, \boldsymbol{u}_{j_r}\}$ for the corresponding eigenspace give the same sum $|\boldsymbol{u}_{j_1}\rangle \langle\boldsymbol{u}_{j_1}| + \cdots + |\boldsymbol{u}_{j_r}\rangle \langle\boldsymbol{u}_{j_r}|$.

Proof. If there is just one distinct eigenvalue λ, then $\boldsymbol{A}$ is multiplication by λ. Then λ being real and the uniqueness of the sum (14.1) are clear. So suppose the eigenspace W of λ_1 is not the whole space. Now consider any vectors $\boldsymbol{x}$ in W and $\boldsymbol{y}$ in the orthogonal complement $W^\perp$ of W. Because $\boldsymbol{A}$ is Hermitian,

$$\langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{y}\rangle = \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}\rangle = \lambda_1^* \langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0.$$

Thus $W^\perp$, as well as W, is closed under $\boldsymbol{A}$, so we may consider the restrictions of $\boldsymbol{A}$ to each space. Since they have lower dimension and each has fewer distinct eigenvalues, the proof follows by induction—in particular, that all the eigenvalues are real. □

For an example of using the spectral theorem, we compute square roots of the Pauli matrix

$$\boldsymbol{Y} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}.$$

The eigenvalues are $+1$ with eigenvector $\phi_+ = \frac{1}{\sqrt{2}}(1, i)$ and -1 with eigenvector $\phi_- = \frac{1}{\sqrt{2}}(1, -i)$. Recall that complex numbers inside $\langle\cdot|$ are conjugated. We can first verify

$$(+1)\,|\phi_+\rangle \langle\phi_+| + (-1)\,|\phi_-\rangle \langle\phi_-| = \frac{1}{2}\left(\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix} - \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}\right) = \boldsymbol{Y}.$$

Now we can get a square root by using any square roots of the eigenvalues. The choice $1^2 = 1$ and $i^2 = -1$ gives

$$\boldsymbol{Y}^{1/2} = \frac{1}{2}\left(\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix} + i\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} 1+i & -1-i \\ 1+i & 1+i \end{bmatrix}.$$

Note that we can take out a factor of $\sqrt{i} = e^{i\pi/4} = \frac{1+i}{\sqrt{2}}$ and what is left is

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

which is like the Hadamard matrix rotated 90 degrees left. Its transpose is likewise a square root and is also sometimes labeled "$\boldsymbol{Y}^{1/2}$"—note also that while $\boldsymbol{Y}$ is Hermitian, none of its square roots is Hermitian.

14.6.1 Trace and POVM

The **trace** $\mathrm{Tr}(\rho)$ of ρ, defined as the sum of its diagonal entries, is $\sum_i p_i = 1$. To apply a unitary linear transformation $\boldsymbol{A}$ to ρ is not simply to multiply $\boldsymbol{A}$ through the sum as $\sum_i p_i \boldsymbol{A}\,|\phi_i\rangle\,\langle\phi_i|$ but, rather, to apply the so-called double action:

$$\boldsymbol{A}\rho\boldsymbol{A}^* = \sum_i p_i \boldsymbol{A}\,|\phi_i\rangle\,\langle\phi_i|\,\boldsymbol{A}^* = \sum_i p_i\,|\boldsymbol{A}\phi_i\rangle\,\langle\boldsymbol{A}\phi_i|$$

Because $\boldsymbol{A}$ is unitary, the double action preserves the trace.

The definition of measurement that is considered canonical for mixed states is similar to definition 14.3.

DEFINITION 14.6 A **positive operator valued measure** (**POVM**) is given by a set $\{\boldsymbol{E}_1, \ldots, \boldsymbol{E}_m\}$ of PSD matrices such that

$$\sum_{j=1}^{m} \boldsymbol{E}_j^* \boldsymbol{E}_j = \boldsymbol{I}.$$

The probability of outcome j on measuring a mixed state $\rho = \sum_i p_i\,|\phi_i\rangle\,\langle\phi_i|$ is

$$p_j = \mathrm{Tr}(E_j\rho),$$

and if a matrix $\boldsymbol{M}_j$ is known such that $\boldsymbol{E}_j = \boldsymbol{M}_j^* \boldsymbol{M}_j$, then the next state is

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

The last point contains a subtlety. By $\boldsymbol{E}_j$ being PSD there exists a matrix $\boldsymbol{M}_j$ such that $\boldsymbol{E}_j = \boldsymbol{M}_j^* \boldsymbol{M}_j$, but it is not unique. The matrices $\boldsymbol{M}_j$ need to be specified in order to engineer the measurement for continued operations with ρ' but if only the act of sampling outcomes according to the distribution $\{p_j\}$ is needed, then they can be dispensed with in the analysis. The denominator is p_j not $\sqrt{p_j}$ because the outer products used in ρ already multiply pairs of amplitudes. The index j is not the same as i: the POVM may have many more (or fewer) components than the representation of the mixed state.

Recall from the Bloch sphere that a mixed state of a single qubit—no matter how many pure states were averaged to produce it—has a canonical representation as a binary distribution of two pure states. All of its representations give the same ρ, and it follows from the above that *there is no quantum experiment that can distinguish those representations*. The ρ is all we know.

Returning to our example, we can now compute what Alice will measure. If her qubit is really the mixture

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix},$$

then after applying the Hadamard gate she will have

$$\begin{aligned} \rho' &= \boldsymbol{H}\rho\boldsymbol{H} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Alice's final standard-basis measurement then gives:

$$|0\rangle \quad \text{with probability} \quad \langle 0|\ \rho'\ |0\rangle = \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 0.5;$$

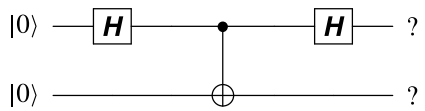
$$|1\rangle \quad \text{with probability} \quad \langle 1|\ \rho'\ |1\rangle = \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0.5.$$

So Alice will be disappointed—her final qubit is not $|0\rangle$, and she will fail to get that outcome half the time.

14.6.2 Partial Traces

Now let us unfold a different scenario involving only pure states. Bob visited the store just before Alice and bought a few $|0\rangle$ qubits to take on a trip to Alpha

Centauri. Unknown to anyone, one of them got entangled with the qubit later sold to Alice via a silent ***CNOT*** operation. Occurrences like this are common in our world. When it happens to the vendor's spare-sample qubits, he is measuring half of a Bell pair $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, which still gives the 50-50 statistics he expects. But even without Bob measuring his entangled qubit, Alice does not get the $|0\rangle$ result she expects, because when she takes her qubit home and applies ***H***, she finds herself measuring the first qubit of the following two-qubit circuit:



The ***CNOT*** affects the first qubit so that the two ***H*** gates on do *not* cancel. Instead, the final $\boldsymbol{H} \otimes \boldsymbol{I}$ maps the Bell pair to

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & -1 \end{bmatrix}.$$

This state $|\kappa\rangle$ remains entangled, but Alice does not always get $|0\rangle$ when she measures its first qubit in the standard basis. The projectors for this are

$$\{P_0, P_1\} = \left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right\}.$$

For outcome $|0\rangle$ the probability is

$$\langle\kappa|\ P_0\ |\kappa\rangle = \frac{1}{4}\langle(1,1,0,0),(1,1,1,-1)\rangle = \frac{1}{2}.$$

So Alice is equally disappointed in this scenario. Note, incidentally, that this determination is irrespective of whether Bob measures his qubit. If he did and got $|0\rangle$, then it means that Alice walked out of the store with $|0\rangle$ (not $|+\rangle$ as she believed) and that her applying ***H*** changed it to $|+\rangle$. If Bob got $|1\rangle$, then Alice will have $\boldsymbol{H}|1\rangle = |-\rangle$, but again, her measurement in the standard basis will give 50-50 odds.

There are other ways that Alice's distribution of results would hinge on what Bob saw, but since there is no communication from Bob—he could be near Alpha Centauri before he examines his qubit—it does not affect Alice's understanding of her distribution. What is remarkable is that quantum mechanics

provides a characterization of Alice's distribution that paints Bob out of the picture, without needing to care what he does.

The relevant operation—the last quantum primitive we treat here—is the **partial trace**. If we have an underlying product Hilbert space $\mathbb{H} = \mathbb{X} \otimes \mathbb{Y}$ of dimension $d_1 \cdot d_2$ and a tensor product $\boldsymbol{U} = \boldsymbol{V} \otimes \boldsymbol{W}$, then the partial trace $\mathrm{Tr}_{\mathbb{Y}}$ **over** $\mathbb{Y}$ maps $\boldsymbol{U}$ to $\boldsymbol{V}\mathrm{Tr}_{\mathbb{Y}}\boldsymbol{W}$. Because such products linearly span the space of all (unitary) linear operators on $\mathbb{H}$, we have uniquely specified $\mathrm{Tr}_{\mathbb{Y}}$ in a basis-invariant manner. Relative to the standard basis, the trick is to tile a $d_1 d_2 \times d_1 d_2$ matrix $\boldsymbol{M}$ with copies of the d_2-dimensional identity matrix, add up the entries of $\boldsymbol{M}$ that fall on the diagonal of each copy, and output the resulting $d_1 \times d_1$ matrix as $\mathrm{Tr}_{\mathbb{Y}}(M)$.

Let us now trace out Bob from the pure state κ. We first form

$$|\kappa\rangle\langle\kappa| = \frac{1}{4}\begin{bmatrix} \mathbf{1} & 1 & \mathbf{1} & -1 \\ 1 & \mathbf{1} & 1 & \mathbf{-1} \\ \mathbf{1} & 1 & \mathbf{1} & -1 \\ -1 & \mathbf{-1} & -1 & \mathbf{1} \end{bmatrix},$$

where we have bolded the entries that are added pairwise. The resulting 2×2 matrix is

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

This is a density matrix. It is the same as the density matrix we computed for Alice in the first scenario. This is not an accident.

THEOREM 14.7 For every mixed state ρ of n qubits, we can build a pure state $|\kappa\rangle$ of $2n$ qubits such that $\rho = \mathrm{Tr}_{\mathbb{Y}}(|\kappa\rangle\langle\kappa|)$, where $\mathbb{Y}$ is the space of the added qubits.

To prove this, use theorem 14.5 to write ρ as a sum $\sum_i \lambda_i |\phi_i\rangle\langle\phi_i|$, where the $|\phi_i\rangle$ form an orthonormal basis of the space $\mathbb{X}$ of the original n qubits. Then, taking $\mathbb{Y}$ to be another copy of $\mathbb{X}$, we can define

$$\kappa = \sum_i \sqrt{\lambda_i}\, |\phi_i\rangle |\phi_i\rangle .$$

Verifying $\rho = \mathrm{Tr}_{\mathbb{Y}}(|\kappa\rangle\langle\kappa|)$ is exercise 14.9. The upshots of all this are as follows:

- Every mixed state is potentially the trace-out of a pure state.

- The holder of one part of an entangled system cannot observe any change to her system resulting from actions by other parties on their parts, as long as there is no other communication among them.

The combination of these two allows us to treat mixed states as complete known entities rather than incomplete results from lack of knowledge. When further interactions connect the parties, however, differences resulting from their actions can be observed. The next section gives a prime example.

The main significance for us now is how entanglement with Bob destroyed coherence from Alice's point of view. Specifically, it prevented the interference that makes two consecutive Hadamard gates cancel. The result is that, with some frequency, Alice will experience a wrong value from her qubit.

14.6.3 Depolarizing and Dephasing

Errors in a single qubit during the course of a computation can be modeled as unwanted interactions with the environment. The analysis has informative symmetry when the errors of a **bit flip** (i.e., multiplication by $\boldsymbol{X}$), **phase flip** (multiplication by $\boldsymbol{Z}$), or both (multiplication by $\boldsymbol{Y}$ ignoring global phase) are considered equally likely with probability $\frac{p}{3}$. With reference to the Bell basis in section 14.5, the action on the entangled pair $|\Phi^+\rangle$ is given by the density matrix evolution $|\Phi^+\rangle\langle\Phi^+| \mapsto \rho'$ where

$$\begin{aligned}\rho' &= (1-p)\,|\Phi^+\rangle\langle\Phi^+| + \frac{p}{3}(|\Psi^+\rangle\langle\Psi^+| + |\Phi^-\rangle\langle\Phi^-| + |\Psi^-\rangle\langle\Psi^-|)\\ &= (1-p')\,|\Phi^+\rangle\langle\Phi^+| + p'\Big(\frac{\boldsymbol{I}}{4}\Big),\end{aligned}$$

where $p' = \frac{4}{3}p$ and $\frac{\boldsymbol{I}}{4}$ is the density matrix for the useless point at the center of the Bloch sphere for the two-qubit system. This presumes that $p \le \frac{3}{4}$; note that $p = \frac{3}{4}$ produces the complete mixture of the Bell basis. The **fidelity** of ρ' to the original state is given by

$$F = \langle\Phi^+|\ \rho'\ |\Phi^+\rangle = 1 - p.$$

With reference to Alice's Bloch sphere, this action contracts any point inward by a factor of p'. Since the Bloch sphere is typified by spin polarization, the action is called **depolarization**.

There is a second basic way that Alice's qubit can lose fidelity. Suppose she buys both a $|+\rangle$ and a $|-\rangle$ and asks for each to be supplied as a point on a Bloch sphere. Customer service at the back of the store puts the spheres

in a special shielding bag for Alice to take to checkout. Alas, the path to the cash register goes through the sections selling beam generators and measuring devices, which are crowded with many Bobs. As Alice pushes past them they jostle her bag so that her qubits jiggle on their spheres. The pure $|+\rangle$ state becomes the average of all the jiggles—which is a point *inside* the sphere, hence a properly mixed state. If the first jiggle is a rotation by θ around the z-axis of its Bloch sphere, then the original density matrix

$$\rho_+ = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

becomes

$$\rho'_+ = \frac{1}{2}\begin{bmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & e^{i\theta} \\ e^{-i\theta}1 & 1 \end{bmatrix}.$$

The effect on her $|-\rangle$ is similar. When the bobbing about the z-axis is modeled as a Gaussian scattering process with variance 4λ, the expected values of her mixed states become

$$\rho''_+ = \frac{1}{2}\begin{bmatrix} 1 & e^{-\lambda} \\ e^{-\lambda}1 & 1 \end{bmatrix}, \qquad \rho''_- = \frac{1}{2}\begin{bmatrix} 1 & -e^{-\lambda} \\ -e^{-\lambda}1 & 1 \end{bmatrix}.$$

The more this happens, the more the off-diagonal elements tend toward zero. This happens equally to ρ_+ and ρ_-, losing the distinction between them. If it goes all the way to zero, then she is left with $\rho_0 = \frac{1}{2}\boldsymbol{I}$ in both cases, and any further operation $\boldsymbol{U}\rho_0\boldsymbol{U}^* = \frac{1}{2}\boldsymbol{U}\boldsymbol{U}^* = \rho_0$ has no effect. In general, we can represent the effect on a state ρ as a mixture

$$F\rho + (1 - F)\boldsymbol{D},$$

where $\boldsymbol{D}$ is a diagonal matrix of unit trace and again F, $0 \leq F \leq 1$, is the **fidelity**. The Bloch sphere again contracts, this time toward the z-axis like a deflating American football on its tip. This is **dephasing**.

The bit-flip and phase-flip errors can be targeted by *quantum error-correcting codes* provided the overall error rate p is small enough. This is the import of the **quantum fault tolerance theorem**, whose beautiful theory is beyond our scope in this text. The threshold for p has not been met, and the codes impose substantial overhead, so the current era is one of **noisy intermediate-scale quantum** (NISQ) devices that try to complete useful computations before the fidelity is completely lost.

14.7 The CHSH Game

Alice and Bob enter now on equal footing. They are still incommunicado with each other, but they may have several points of common contact:

- A referee, Ralph, communicates with each of them separately.
- They can have reached prior agreements before going incommunicado. We might also allow them to confer between *trials* that involve communications with Ralph.
- They can observe the same source of classical random bits—that is, they share a coin.
- In the quantum case they may (instead) share an entangled Bell pair $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.

In the **CHSH game**, named for John Clauser, Michael Horne, Abner Shimony, and Richard Holt in the paper Clauser et al. (1969), a *trial* goes as follows:

1. Ralph generates two bits $a, b \in \{0, 1\}$ uniformly at random, sends a to Alice, and sends b to Bob.
2. Alice and Bob each say "yes" or "no" in separate replies to Ralph.
3. Alice and Bob win if their answers differ when $a = b = 1$ and agree otherwise. They play cooperatively.

In the classical case they have a simple way to win 75% of the time: they both always say no. (Or they both always say yes.) They win unless Ralph's two random coins both give 1. The main question is, can they do better?

14.7.1 Classical Case

In the classical setting, the answer is *no*. After observing the shared coins, Alice can have only four different behaviors S_A in any trial:

- **YY**: Say yes regardless of whether $a = 0$ or $a = 1$
- **YN**: Say yes to $a = 0$, no to $a = 1$
- **NY**: Say no to $a = 0$, yes to $a = 1$
- **NN**: Say no to both

Bob has the same options for his behavioral strategy S_B. Bob and Alice can agree on a fixed behavior or on how outcomes of their shared coins map to behaviors for each, so that each may know the behavior the other has elected. As in the case of the 75%-assuring strategy **NN**, one can even know the other's response ahead of time. What cannot be known, however, is the value of the other's *bit* from Ralph. The basic fact is this:

> For any pair $(S_A, S_B) \in \{\mathbf{YY,YN,NY,NN}\}^2$, there is a combination $(a, b) \in \{0, 1\}^2$ that causes Alice and Bob to lose.

Since that combination comes with probability at least 0.25 from Ralph, the success probability cannot be raised above 0.75. No scheme of how each can interpret results from their shared classical randomness changes this. In particular, nor can an oracle Ozzie, controlling the shared "random" bits, guide Alice and Bob to winning responses with any higher effectiveness. We interpret this after covering the quantum case.

14.7.2 Quantum Case

It seems at first that sharing the Bell pair has no more use than sharing the classical random coins. If Alice does a standard measurement of her qubit of the pair and sees $|0\rangle$, this means Bob will also certainly see $|0\rangle$ when he measures, but so what—it is the same with a shared coin, likewise when both see $|1\rangle$. They can do things to bias the outcomes away from 50-50, but the classical Alice and Bob can do the same by mapping results of multiple shared coins differently for each. The subtle difference is that Alice and Bob can rig how their shared qubit behaves in ways that change the odds each experiences.

For intuition, let us revisit the use of polarizing filters as measuring devices. Let us suppose that north-south linear polarization of the entangled qubits means $|1\rangle$ and that east-west means $|0\rangle$. Alice and Bob each have a sheet of polarizing filter that each can orient in his or her chosen direction. They do not have to use the same direction, and each can choose a direction based on the bit from Ralph—this is key. We equip each with detector sensitive enough to register a single photon. (Such detectors are available commercially, though getting 95% or even 90% success rate still comes at a premium.) As viewed by Ralph, we suppose that his messages arrive to Alice and Bob simultaneously, that Alice's shared qubit comes through her filter a designated time interval later (granting enough time for her to choose an orientation), and that Bob's

qubit goes through his filter a microsecond later. Moreover, Alice will say yes if and only if she gets a ping from her detector, and Bob will do likewise. Thus, we can make their responses involuntary so that the whole trial concludes—pending only Ralph receiving and judging the responses—at the instant Bob measures. Since light travels just under 300 meters in a microsecond, spacing Alice and Bob more than that apart ensures that no communication between them can physically occur in the microsecond between their measurements.

Thus, everything about strategy comes down to Alice's and Bob's orientations of their filters. First suppose Alice aligns hers north-south (we'll just say N). Then she will get a ping if the Bell pair comes through as $|11\rangle$, nothing if $|00\rangle$. If she aligns east-west (E), then she always gets a ping on $|00\rangle$, nothing on $|11\rangle$. If she aligns it southwest-northeast (NE), then either basic outcome becomes a 50-50 coin flip on whether it registers a ping. The key physical point about the measurement the filters perform was discussed in section 14.5:

> If Alice gets a ping from her NE axis, the photon becomes polarized NE. Because Bob's qubit is entangled, his photon has the same property. Conversely, when Alice gets no ping, both photons snap to the NW axis.

We will later redo the quantum analysis without relying on this property of physics—and this will in turn explain why the physical behavior happens as mathematical consequence. For now we go to a finer angle and suppose Alice orients her filter axis 22.5 degrees rather than 45 degrees up from east. On the compass this is east-northeast, ENE. If the shared qubits come through as $|00\rangle$, then Alice's chance of getting a ping will be

$$p = \cos^2\left(\frac{\pi}{8}\right) = \frac{\sqrt{1/2}+1}{2} = 0.85355\ldots$$

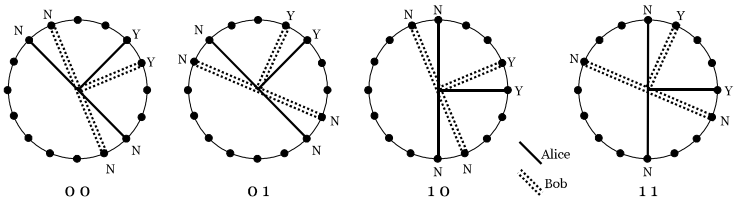
Now picture Bob doing this instead. If Alice holds her filter E and gets a ping, the shared qubits will be (or were or are) in state $|00\rangle$, and Bob will get a ping with the same $0.85355\ldots$ probability. But if Alice gets no ping, then the qubits are $|11\rangle$ and Bob's filter—which is close to orthogonal to the north axis with $|1\rangle$—will have only a $1-p$ chance of letting his photon through. Either way, there is a $0.85355\ldots$ chance that Bob and Alice get the same result, whether "ping" or "no ping."

If Alice instead holds her filter NE, then Bob's angle will be 22.5 degrees "under" rather than "over," but the analysis is much the same. If Alice gets a ping, then Bob's photon becomes NE and his NNE setting again lets it through

with 0.85355... probability, so with that probability his answer will agree. If Alice gets no ping, then the photons are aligned NW (to SE) and Bob's ENE is again almost orthogonal, so with 0.85355... probability he gets no ping. Thus, ENE for Bob makes him likely to agree whether Alice chooses E or NE. But now suppose Bob chooses north-northeast (NNE) instead, while Alice has chosen E. Now their filters are oriented 67.5 degrees part, so when one gets a ping, the other likely does not—so likely their answers will disagree. We want this to happen when and only when Ralph has sent $a = b = 1$. Thus, we have hit on a winning strategy:

- Alice chooses NE if $a = 0$, else she chooses E.
- Bob chooses ENE if he gets $b = 0$, else he chooses NNE.

Figure 14.6
Basis-choice strategy for Alice and Bob in CHSH game.



If Alice gets $a = 0$, her NE axis is at 22.5 degrees from both options for Ralph, so with probability 0.85355... their answers agree and they win. If Bob gets $b = 0$, then his ENE axis is likewise close to both of Alice's options, so again with probability 0.85355... their answers agree which is what they need to win in this case too. But if Ralph sends 1 to both, their axes are E and NNE respectively, 67.5 degrees, so with probability 0.85355... their answers are different, which is what they then need to win. This is shown in figure 14.6.

So their theoretical winning probability in all cases is 0.85355.... Even if their detectors each err on 6% of the photons (independently), they will still win on more than $0.88 * 0.85355 = 0.751...$ of the trials, which beats the 75% classical maximum. In fact, physical runs of this game using highly accurate detectors have produced a win rate in valid trials of over 84%.

For a second look at the quantum case, we will put away the polarizing sheets and stay in the standard basis for measurements. The only "snap" or

"collapse" will be to $|00\rangle$ or $|11\rangle$, but this will not give us anything new to think about because we've already granted that those are the possibilities.

What Alice and Bob choose to do instead is apply different 2×2 operators $\boldsymbol{A}$ and $\boldsymbol{B}$ to their halves of the pair. Of course, these will be the same change-of-basis operators as in our intuitive description, but it will be fun to see the workings. The general rotation by $+\theta$ in the plane with $|0\rangle$ at $(1,0)$ (i.e., "east") and $|1\rangle$ at $(0,1)$ is

$$R_y(2\theta) = \cos(\theta)\boldsymbol{I} - i\sin(\theta)\boldsymbol{Y} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

The "2θ" is because in the Bloch sphere this is a rotation of 2θ (clockwise) around the y-axis. Without further ado, here are the strategies:

- Alice applies $R_y\left(\frac{\pi}{2}\right)$ to her qubit if she gets $a = 0$ from Ralph, else she does nothing.
- Bob applies $R_y\left(\frac{\pi}{4}\right)$ to his qubit if he gets $b = 0$, else he applies $R_y\left(\frac{3\pi}{4}\right)$.

Both then measure in the standard basis and say no on $|0\rangle$, yes on $|1\rangle$. If Alice gets $a = 1$, then the two computations resulting from Bob's actions are

$$b{=}0: \begin{bmatrix} \cos(\frac{\pi}{8}) & -\sin(\frac{\pi}{8}) & 0 & 0 \\ \sin(\frac{\pi}{8}) & \cos(\frac{\pi}{8}) & 0 & 0 \\ 0 & 0 & \cos(\frac{\pi}{8}) & -\sin(\frac{\pi}{8}) \\ 0 & 0 & \sin(\frac{\pi}{8}) & \cos(\frac{\pi}{8}) \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos(\frac{\pi}{8}) \\ \sin(\frac{\pi}{8}) \\ -\sin(\frac{\pi}{8}) \\ \cos(\frac{\pi}{8}) \end{bmatrix}.$$

$$b{=}1: \begin{bmatrix} \cos(\frac{3\pi}{8}) & -\sin(\frac{3\pi}{8}) & 0 & 0 \\ \sin(\frac{3\pi}{8}) & \cos(\frac{3\pi}{8}) & 0 & 0 \\ 0 & 0 & \cos(\frac{3\pi}{8}) & -\sin(\frac{3\pi}{8}) \\ 0 & 0 & \sin(\frac{3\pi}{8}) & \cos(\frac{3\pi}{8}) \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos(\frac{3\pi}{8}) \\ \sin(\frac{3\pi}{8}) \\ -\sin(\frac{3\pi}{8}) \\ \cos(\frac{3\pi}{8}) \end{bmatrix}.$$

When $b = 0$ the outcomes agree with probability $\cos^2\left(\frac{\pi}{8}\right) = 0.85355\ldots$ as before. When $b = 1$, bringing out the case of both getting 1 from Ralph, their joint result is $|01\rangle$ or $|10\rangle$ with that probability, whereupon their answers disagree and they win. When $a = 0$ the math is a little different. Alice applies

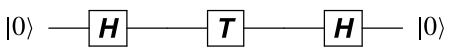
$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

Bob then applies the same matrices as above, getting:

$$b=0: \frac{1}{2}\begin{bmatrix} \cos(\frac{\pi}{8})+\sin(\frac{\pi}{8}) \\ \sin(\frac{\pi}{8})-\cos(\frac{\pi}{8}) \\ \cos(\frac{\pi}{8})-\sin(\frac{\pi}{8}) \\ \sin(\frac{\pi}{8})+\cos(\frac{\pi}{8}) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \sqrt{2}\cos(\frac{\pi}{8}-\frac{\pi}{4}) \\ \sqrt{2}\sin(\frac{\pi}{8}-\frac{\pi}{4}) \\ \sqrt{2}\sin(\frac{\pi}{4}-\frac{\pi}{8}) \\ \sqrt{2}\cos(\frac{\pi}{8}-\frac{\pi}{4}) \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} \cos(\frac{-\pi}{8}) \\ \sin(\frac{-\pi}{8}) \\ \sin(\frac{\pi}{8}) \\ \cos(\frac{-\pi}{8}) \end{bmatrix},$$

where we used Ptolemy's angle identities to replace the sums. The rest is the same as before, and when $b=1$, the symmetry of the sums and differences under replacing $\frac{\pi}{8}$ by $\frac{3\pi}{8}$ gives the same result: probability $0.85355\ldots$ that Alice and Bob will win by their answers agreeing.

A corollary of this is that the quantum probabilities cannot be modeled by classical random coins. This is easier to see in one sense: the success probability $0.85355\ldots$ is not a rational number. Note that preparing the Bell pair used only one Hadamard gate for nondeterminism. Alice's operator $R_y\left(\frac{\pi}{2}\right)$, when she uses it, is equivalent to one of the square roots of $\boldsymbol{Y}$ found in section 14.6, which has the same entries as $\boldsymbol{H}$. Bob's operator is not quite the same as the sequence $\boldsymbol{HTH}$ (the difference is explored in the exercises) but we can convey the essential point by referring to it. The amplitude of



is $\alpha=\frac{1}{2}(1+\omega)$ where $\omega=e^{i\pi/4}$. The probability is

$$|\alpha|^2=\cos^2\left(\frac{\pi}{8}\right)=0.85355\ldots$$

again. If the two $\boldsymbol{H}$ gates could be modeled by classical random coins, the most extreme probability less than 1 they could give would be 0.75. If they gave rise to a larger finite number of classical fair coins, the denominator of α would be a power of 2. It is nevertheless possible to analyze the logic of the quantum circuit using one binary classical variable for each $\boldsymbol{H}$ gate, as we will remark in chapter 19.

14.8 Quantum Supremacy

The CHSH game represents one kind of quantum supremacy. It is provable and has been demonstrated to high precision. It is however for an interactively-defined problem rather than a straight-up computational task. It does "scale up" to cases of many Alice-Bob interactions and multiple parties and is emblematic

of important applications in quantum communication theory. We have seen computational tasks where quantum algorithms beat classical ones, but where one of the following is true:

- they do not scale up to higher input sizes n;
- they scale up but are defined with regard to an oracle function F; or
- they scale up without an oracle function, but their classical intractability has not been proven, nor has there been a convincing demonstration of quantum advantage that scales up.

Hence, the search for a demonstration that overcomes these objections has ramped up in recent years. In October 2019 a supremacy claim was made by a team of researchers led by Google (Alphabet, Inc.) and the University of California at Santa Barbara (see Arute et al., 2019). The principle is elementary enough to cover here and also illustrates a few discussion points in this chapter.

Every quantum circuit C with n qubit lines induces a probability distribution D_C over $z \in \{0,1\}^n$ by

$$D_C(z) = |\langle z|\ C\ |0^n\rangle|^2.$$

We could postulate extra ancilla qubit lines and make the input $|0^m\rangle$ with $m > n$, but the reported experiment does not do so. Since C could have initial ***NOT*** gates on some lines, it does not matter that the *input* is fixed as $|0^n\rangle$ rather than sampled. Sampling the *output* is the game.

Next we consider probability distributions D' that are generated uniformly at random by the following process:

> For $i = 1$ to $R = 2^r$:
> (a) choose a $z \in \{0,1\}^n$ uniformly at random; then
> (b) increment its probability $D'(z)$ by $\frac{1}{R}$.

Here we intend r to be the number of binary nondeterministic gates in the circuit. In place of Hadamard the circuits actually use the $\boldsymbol{V}$-gate, that is, the square root of ***NOT***, which is also written $\sqrt{\boldsymbol{X}}$ or $\boldsymbol{X}^{1/2}$, plus the gate $\boldsymbol{Y}^{1/2}$ in section 14.6 and the gate $\boldsymbol{W}^{1/2}$ in problem 14.7 below. The difference matters to technical analysis of the distributions D_C, but the interplay between quantum nondeterministic gates and classical random coins remains in force. Now we can state the task first in general terms and then more particularly:

> Given randomly generated quantum circuits C as inputs, distinguish D_C with high probability from any D'.

The specific task involves a number $\delta > 0$ and moderately large integer k that are set by the terms of the experiment in order to be technologically feasible for quantum devices even amid decoherence errors and yet concretely infeasible for classical computation.

> Given randomly generated C, generate samples $z_1, \ldots, z_k \in \{0,1\}^n$ such that $\frac{1}{k}(D_C(z_1) + \cdots D_C(z_k)) \geq 1 + \delta$.

A baseball analogy may help, in which we take $r = n$, so $R = N = 2^n$, to sharpen differences. We are distributing N units of probability among N "batters" $z \in \{0,1\}^n$. A batter who gets two units hits a double, three units makes a triple, and so on. The key distinction is between the familiar batting average and the *slugging average*, which averages all the bases scored with hits:

- The chance of making an out—that is, getting no units—is $(\frac{N-1}{N})^N$, which is approximately $\frac{1}{e} = 0.367879\ldots$.
- The chance of hitting a single is also about $\frac{1}{e}$, leaving $1 - \frac{2}{e}$ as the frequency of getting an extra-base hit—which makes z a "heavy hitter."
- From k batters chosen uniformly at random, their expected batting average will be $1 - \frac{1}{e} = 0.632\ldots$.
- Their expected slugging average, however, will just be 1: they expect k units to be distributed among them.

Thus, with respect to a random D', and without any knowledge of D', a chosen team of k hitters cannot expect to have a joint slugging average higher than 1. Moreover, for any fixed $\delta > 0$, the chance of getting a slugging average higher than $1 + \delta$ tails away exponentially in k (provided N also grows).

With respect to D_C, however, a quantum device can do better. Google's device programs itself given C as the blueprint, so it just executes C and measures all qubits to sample the output. Finding its own heavy hitters is what a quantum circuit is good at. The probability of getting a hitter who hits a triple is magnified by 3 compared to a uniform choice. Moreover, C will never output a string with zero hits—a "can't miss" property denied to a classical reader of C. For large N the probability distribution approaches xe^{-x} and the slugging

expectation is approximately

$$\int_0^\infty x^2 e^{-x} = -x^2 e^{-x}\Big]_{x=0}^{x=\infty} + 2\int_0^\infty xe^{-x} = 0 - 2xe^{-x}\Big]_{x=0}^{x=\infty} + 2\int_0^\infty e^{-x} = 2.$$

That is, a team $z_1, \ldots, z_k$ drafted by sampling from random quantum circuits C expects to have a slugging average near 2. If C works perfectly, the average will surpass $1 + \delta$ whenever $0 < \delta < 1$ with near certainty as N grows.

The practical challenge is that the implementation of C is not perfect. The consequence of an error in the final output is severe. The heavy-hitter outputs z of a random C are generally not bit-wise similar. Suppose the imperfect C' outputs z' which differs in just one bit from a true output z of C. Then z' is a random sample of n strings at distance 1 from z, but with regard to the true D_C the expected slugging weight reverts to being near 1. Moreover, joint distributions of (C, v) with a large error (bit flip or half-circle phase flip) at point v lose the "can't miss" property, because a z_v output by C for one v may have zero probability of being output by C with alternate error v'. Related phenomena are actually observed physically in disturbances of coherent "speckle" patterns of laser light.

Google's circuits have up to $r = 20n$, so $R \gg N$. Then the "can't miss" aspect of the quantum advantage is less sharp, but the xe^{-x} approximation is closer. By the randomness and scale of their circuits, they can model the effect as a simple loss of fidelity F as represented in section 14.6.3, so that they effectively sample from the distribution

$$F|\langle z|\ C\ |0^n\rangle|^2 + (1 - F)\frac{1}{N}. \tag{14.2}$$

Available reports say that their fidelity is driven below 0.01 but stays above 0.001 in trials. This bounds the range of their experimentally realized δ separation. That it is separated from zero is, however, highly significant. The technology that enables setting δ and framing the size $n = 53$ qubits with 20 layers of binary gates and randomly chosen unary gates is the first of three planks in proving quantum supremacy by this means.

The second plank is to *verify* that outputs $(z_1, \ldots, z_k)$ from the imperfect simulation really do have a slugging average higher than $1 + \delta$ with respect to the true circuits C_i that were presented. This part is done classically, by computing a statistical test that with high probability can be passed only by sets with sufficiently many heavy hitters z_j. The computation for each z_j is expensive, and this is why the separation needs to succeed with a moderate value of k. The point that must be met is that with z_j in hand the test is feasible,

but a classically random search would take too long to amass enough such z_j that could pass the test.

The third—and most contentious—plank is to demonstrate that no feasible classical computation can sufficiently improve on the classical random analysis to pass the tests with more than negligible probability. The sensitive point is that classical algorithms A must be allowed to inspect the blueprints of the randomly generated quantum circuits C as "white boxes." A key piece of theoretical evidence is that problems related to finding heavy-hitting strings are asymptotically hard in terms of average-case complexity, not merely in worst-case complexity. The evidence in question is that the ability of a circuit with classical random coins to sample from close approximations to the distribution (14.2) yields a so-called *Arthur-Merlin protocol* of a kind believed impossible for a certain level of asymptotically hard problems. There remains the issue of concrete hardness. The principle behind the asymptotic hardness and worst-to-average case reduction does arguably take root by $n = 53$, but their concrete tests reduced the number of qubits and/or levels, and the largest ones were dedicated to verifying modeling assertions underlying the second plank. It may be relevant to try exhaustive generation of classical codes A that are small classical circuits or have small specifications. A counter-claim by Pednault et al. (2019) argues the ability of classical hardware to solve the full $n = 53$, $r = 20$ instances within a few days.

If all three planks are established, then we will have a lexically defined problem (given the blueprint of C, find a team of heavy hitters) having a quick quantum solution but no feasible classical solution. It can still be objected that the quantum problem is navel-gazing since it is about quantum circuits C and involves sampling measured outputs in its definition. It is, however, a legitimate *search* problem with a decision variant: given C and a string z_0 of length $< n$, can z_0 be extended to a heavy-hitting string? It appears to have applications beyond the quantum domain that involve certifying randomness for cryptographic purposes. What we emphasize to bring this chapter full circle is that the separation is being demonstrated not just in theory but with a physical device.

14.9 Problems

14.1. Show that the states $\frac{1}{2}(|000\rangle + |100\rangle + |010\rangle + |111\rangle)$ and $\frac{1}{\sqrt{3}}(|00\rangle + |10\rangle + |01\rangle)$ in section 14.3 are entangled.

14.2. Given a 2×2 unitary matrix $\boldsymbol{U}$, prove that the action on the Bloch sphere is a rotation of some number φ degrees around some axis ϕ through the center.

14.3. Suppose $\phi_1, \ldots, \phi_m$ are such that $\sum_i |\phi_i\rangle \langle\phi_i| = \boldsymbol{I}$. Given a unitary matrix $\boldsymbol{A}$, define $\psi_i = \boldsymbol{A}\phi_i$ for each i. Show that $\sum_i |\psi_i\rangle \langle\psi_i| = \boldsymbol{I}$.

14.4. Verify the statement in section 14.4 that the point $\theta = \frac{\pi}{4}$, $\varphi = 0$ is unchanged by applying $\boldsymbol{H}$ to it.

14.5. Show that for every 2×2 unitary matrix $\boldsymbol{U}$ there are real numbers $\theta, \alpha, \beta, \delta$ such that

$$\boldsymbol{U} = e^{i\delta} \boldsymbol{T}_\alpha \boldsymbol{R}_\theta \boldsymbol{T}_\beta.$$

Thus, every 2×2 unitary operation can be decomposed into a rotation flanked by two twists, multiplied by an arbitrary phase shift by δ. Write out the decomposition for the matrix $\boldsymbol{V}$ in problem 3.9. (It does not matter which definition of $\boldsymbol{T}_\alpha$ you use from problem 3.10.)

14.6. Using the measurement formalism in section 14.5 (especially definition 14.3), verify the measurement outcomes when the third qubit in the circuit at the end of section 14.3 is measured, including the resulting state when $|0\rangle$ is observed and the state when $|1\rangle$ is observed.

14.7. Use the spectral method in section 14.6 to calculate a square root of the matrix

$$\boldsymbol{W} = \frac{1}{\sqrt{2}}(\boldsymbol{X} + \boldsymbol{Y}) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1-i \\ 1+i & 0 \end{bmatrix}.$$

14.8. Verify that if $\boldsymbol{A}$ is unitary, and

$$\rho' = \boldsymbol{A}\rho\boldsymbol{A}^* = \sum_i p_i \boldsymbol{A} |\phi_i\rangle \langle\phi_i| \boldsymbol{A}^* = \sum_i |\boldsymbol{A}\phi_i\rangle \langle\boldsymbol{A}\phi_i|,$$

then $\mathrm{Tr}(\rho') = \mathrm{Tr}(\rho)$.

14.9. Verify the statement after theorem 14.7 that $\rho = \mathrm{Tr}_{\mathbb{Y}}(|\kappa\rangle \langle\kappa|)$.

14.10. With reference to the first description of the quantum case of the CHSH game in section 14.7, can Bob win by using an axis pointed NNW as an option? (Does he thereby emulate a famous quotation from *Hamlet* in the form: "I am but mad north-northwest. When the axis is southerly I can tell $|0\rangle$ from $|1\rangle$"?)

14.11. First calculate $\boldsymbol{B} = \boldsymbol{HTH}$ and note how it differs from the real rotation matrix

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

where $\theta = \pi/4$. Then find a 2×2 operator $\boldsymbol{A}$ such that Alice and Bob can win the CHSH game with Alice using the identity and $\boldsymbol{A}$ in her two cases, and Bob using $\boldsymbol{B}$ and a related $\boldsymbol{B}'$ in his two cases.

14.10 Summary and Notes

Besides the claim of achieving quantum supremacy treated in section 14.8, the year 2019 saw renewed discussion in major physics blogs and even the *New York Times* about the nature of quantum mechanics. Some say that the major differences in interpretation, especially of the physical processes accompanying measurements, are immaterial because no experiments are known that can possibly separate them. Others, however, have contended that the interpretation matters to how one conceptualizes quantum *computation*. In this chapter we have tried to avoid any one of these positions. We can characterize our position as a simple realist one: quantum observables and the Born rule for their behavior are just what nature gives us.

We have shown physical ways that quantum mechanics differs from classical mechanics, including the behavior of polarizing filters and the CHSH game. The discussion of depolarization and dephasing follows Preskill (2015) and McDonald (2017), which also have a more precise coverage of the postulates of quantum mechanics. Our coverage of the postulate of how quantum systems evolve is deferred to chapter 18. Also looking ahead to chapter 18, we have addressed notational issues in regard to representing vectors in Dirac notation. Regarding the seeming lack of consensus on notation for outerproducts, the use of $\boldsymbol{u} \otimes \boldsymbol{v}$ for this, clashing with tensor products, comes from Wikipedia's current article on outer products:

https://en.wikipedia.org/wiki/Outer_product

The first description of the CHSH game draws on O'Donnell (2018). The supremacy experiment described in section 14.8 exemplifies how noise in current technology drives the fidelity down. As discussed by Preskill (2018), the advanced algorithms covered here require containing this noise to implement. The supremacy experiment of Arute et al. (2019) grew out of Boixo et al. (2018), Neill et al. (2018), and Villalonga et al. (2019), with input from Aaronson and Chen (2017), Bouland et al. (2018), Markov et al. (2018), and Huang

et al. (2018); see Pednault et al. (2019) for a counterclaim and also Aaronson and Arkhipov (2011), Bremner et al. (2010), Bremner et al. (2016), and Harrow and Montanaro (2017).

We have given more detail on the CHSH game than is typical of secondary sources, including attention to special relativity. This gives some extra support for possible further discussions. One direction is to discuss possible "loopholes" in Bell's-theorem experiments, such as that of Aspect et al. (1982), and efforts to close them, especially Hensen et al. (2015). The 84% success figure is from the former, but the latter needed to throw away more malfunctioning trials. Another direction is to go into the recent extended paradox of Frauchiger and Renner (2018). We pick up a third direction in section 19.7, at the end of chapter 19. Here are some remarks that may help these directions and supplement the understanding from this chapter:

The two CHSH scenarios we described ought to be equivalent, but there are some subtle differences. In the former scenario the polarizing filters effect the measurements; the detectors register only the results. In the latter, Alice and Bob apply the corresponding change-of-basis operators before measuring. The detail in the former that the encounter of Alice's photon with the film occurs an instant before Bob's (at least in Ralph's reference frame) implies that the final state of the qubits will be along one of Bob's axes, pointed ENE or NNE if he gets a ping, or NNW or WNW if he does not. If Bob measured first, then the final state would be along one of Alice's axes: N-S, E-W, NE-SW, or NW-SE. This might last only until the photons fly into the detectors, but it is concrete enough to witness that measurement operators need not commute. In the latter scenario, however, the results were obtained without specifying an order of measuring. Nor does the order of Alice and Bob applying their operators (which are both *local*) matter, and the measurement outcomes are all in the standard basis regardless of who goes first. Note that even though their qubits remain entangled, Alice and Bob *can* get different measurement results in the standard basis. Between them, all of $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ are possible. If it were always $|00\rangle$ or $|11\rangle$, their answers would always agree and they would lose when Ralph sends both 1. In the case where Alice applies $R_y\left(\frac{\pi}{2}\right)$ to her qubit, she does instantaneously change the alignment of Bob's qubit, so that both his options will be 22.5 degrees away from the new axis. And Bob's options, if they came first, would have similar effects on the axis of Alice's qubit. The effect may not seem as vivid as in the polarizing-filter scenario, but it is equally present.

This kind of instantaneous effect is what famously disturbed Einstein. Here it is compounded by the apparent difference of Alice or Bob going first by one microsecond, which is a long interval by computer architecture standards but too short for information to travel 300 meters. It must be said, however, that the probabilities and nature of the observations are exactly the same whichever goes first and in both scenarios. The difference in state described between the scenarios cannot be detected. Most significant, both scenarios and both orders yield the same *density matrix* for Alice's view of her qubit and the same density matrix for Bob's view of his. Therefore we cannot say any quantum information, let alone classical information, is exchanged over that time.

There remains the question of whether the behavior of the shared qubits is foreordained in any one trial. This pertains especially to Alice and Bob getting pings (first scenario) or measuring $|0\rangle$ (second scenario), on which their responses to Ralph depend. It is possible that physical factors—which could be represented as variables in equations under a yet-to-be-discovered theory—could first determine Alice's outcome and then instantly force Bob's. What the above subsections prove, however, is that the totality of those variables cannot have been observed by both Alice and Bob before the instants of their measurements. For if they could, then we could equivalently postulate that an oracle controlling their classical shared coins could impart the values of their variables. If the quantum behavior came down to such *local* hidden variables, then there would be an oracle in the classical setting that enables them to win over 85% of the trials, but we have proved there is none. There still can be hidden variables giving a deterministic analysis of the trials, but they cannot all be local to both Bob and Alice.

Finally, regarding section 14.8, the quantum supremacy claim by the Google-led team remains under evaluation. We gave more details of the statistical testing in our *Gödel's Lost Letter* blog article:

rjlipton.wordpress.com/2019/10/27/quantum-supremacy-at-last/

The size of the experiment is being increased moderately so as to preserve the feasibility of its execution while putting classical emulation efforts further out of reach. Very roughly speaking, the tuning-up adds c to the size so that the effort the team must expend is compounded by a factor of c or c^2, but the emulators—apparently—must compound their hardware and/or time by order exponential in c. We surely have not yet heard the "end of the beginning," let alone the end, of arguments over the nature and degree of advantage brought by physical quantum computers.