16 Quantum Computing and Communications

We've seen many ways in which physical laws set profound limits: messages can't travel faster than the speed of light; transistors can't be shrunk below the size of an atom. Conventional scaling of device performance, which has been based on increasing clock speeds and decreasing feature sizes, is already bumping into such fundamental constraints. But these same laws also contain opportunities. The universe offers many more ways to communicate and manipulate information than we presently use, most notably through quantum coherence. Specifying the state of N quantum bits, called *qubits*, requires 2^N coefficients. That's a lot.

Quantum mechanics is essential to explain semiconductor band structure, but the devices this enables perform classical functions. Perfectly good logic families can be based on the flow of a liquid or a gas; this is called *fluidic logic* [Spuhler, 1983] and is used routinely in automobile transmissions. In fact, quantum effects can cause serious problems as VLSI devices are scaled down to finer linewidths, because small gates no longer work the same as their larger counterparts do.

But what happens if the bits themselves are governed by quantum rather than classical laws? Since quantum mechanics is so important, and so strange, a number of early pioneers wondered how a quantum mechanical computer might differ from a classical one [Benioff, 1980; Feynman, 1982; Deutsch, 1985]. Feynman, for example, objected to the exponential cost of simulating a quantum system on a classical computer, and conjectured that a quantum computer could simulate another quantum system efficiently (i.e., with less than exponential resources). We now know this to be true, and it turns out to be just one of the remarkable implications of representing information quantum mechanically. A quantum computer could find prime factors in polynomial time rather than the exponential time required classically, thereby defeating conventional cryptosystems used for electronic commerce and information security. Quantum mechanics also offers an entirely new way to protect information by applying the foundations of quantum measurement rather than number theory, and it can even be used to teleport particles instead of just communicating bits.

To understand the engineering implementation of such science-fiction fantasies it's necessary to understand quantum mechanics (to the extent that's possible). This will entail looking more deeply at the formal structure of the theory than we have in our passing uses of it so far, then extending it to include angular momentum and thermodynamics, and finally building up a language to describe quantum information, protocols, circuits, and algorithms.

16.1 QUANTUM MECHANICS

Of the many attempts to explain the foundations of quantum mechanics, arguably the most successful one is: *because that's the way it is*. Experimental observations have inspired and then justified the development of a theory that is remote from our everyday experience. For this reason, intuition is dangerous in quantum mechanics. At the outset, when in doubt, it's best to focus on the mathematical formalism and leave questions about meaning and interpretation to later when intuition is tempered by experience.

16.1.1 States and Operators

The state of a quantum system is completely specified by its *wave function* $|\psi\rangle$. *Operators* associate transformations of the wave function with observable properties. If an operator $\hat{X}$ acting on a wave function $|n\rangle$ returns the wave function multiplied by a coefficient x_n

$$\hat{X}|n\rangle = x_n|n\rangle \tag{16.1}$$

then $|n\rangle$ is said to be an *eigenvector* of the operator, x_n the corresponding *eigenvalue*, and the index n is the *quantum number* of the state. The spectrum of eigenvalues can be continuous, such as the position of a free particle, or discrete, like the energy levels of an atom.

Quantum operators are linear, which means that

$$\hat{X}\left(\alpha|a\rangle + \beta|b\rangle\right) = \alpha\hat{X}|a\rangle + \beta\hat{X}|b\rangle \quad , \tag{16.2}$$

for arbitrary complex constants α and β. This is an experimental observation, although if quantum mechanics *was* nonlinear then it would be possible to send messages faster than light or solve intractable computational problems [Weinberg, 1989; Abrams & Lloyd, 1998] – remarkable but unlikely capabilities.

A complete set of eigenfunctions forms a basis that can be used to expand an arbitrary wave function

$$|\psi\rangle = \sum_n a_n|n\rangle \quad . \tag{16.3}$$

$|a_n|^2$ is the probability that a measurement of $\hat{X}_n$ will return the value x_n. If this does happen, then the wave function is said to *collapse* into the state $|n\rangle$ regardless of what basis it started in. A quantum measurement necessarily disturbs the system unless it is already in an eigenstate.

The coefficients a_n provide a representation of the wave function, conventionally written as a column vector called the *state vector*. The space of these vectors is called a *Hilbert space*, which will be finite- or infinite-dimensional for discrete or continuous eigenvalues respectively. Choosing a different set of eigenfunctions gives a different representation for the state vector. Operators on state vectors are matrices.

In the state vector representation, quantum mechanical manipulations are just linear algebra [Strang, 1988]. Associated with each column vector $|\psi\rangle$ is a row vector $\langle\psi|$

$$\langle\psi| = |\psi\rangle^\dagger \quad , \tag{16.4}$$

where the dagger operator is defined to be the complex conjugate of the transpose

$$\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}^{\dagger} = \begin{pmatrix} a^* & c^* & e^* \\ b^* & d^* & f^* \end{pmatrix} \tag{16.5}$$

and is called the *Hermitian adjoint*. Taking the Hermitian adjoint of a product reverses the order

$$\left(\hat{X}\hat{Y}\right)^{\dagger} = \hat{Y}^{\dagger}\hat{X}^{\dagger} \tag{16.6}$$

and so in particular

$$\left(\hat{X}|\psi\rangle\right)^{\dagger} = \langle\psi|\hat{X}^{\dagger} \quad . \tag{16.7}$$

The adjoint of a scalar is just the complex conjugate

$$\left(x|\psi\rangle\right)^{\dagger} = \langle\psi|x^* \quad . \tag{16.8}$$

The *inner product* of two state vectors is

$$\begin{aligned} \langle\psi|\varphi\rangle &= (\psi_1^* \ \psi_2^* \ \cdots) \times \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \end{pmatrix} \\ &= \psi_1^*\varphi_1 + \psi_2^*\varphi_2 + \cdots \quad . \end{aligned} \tag{16.9}$$

Dirac named $\langle\psi|$ a *bra*, $|\varphi\rangle$ a *ket*, and $\langle\psi|\varphi\rangle$ a ... *bracket* (get it?).

Eigenfunctions are orthonormal,

$$\langle j|k\rangle = \delta_{ij} \quad . \tag{16.10}$$

Wave functions are therefore normalized because the sum of the probabilities of measurement outcomes must equal 1:

$$\begin{aligned} \langle\psi|\psi\rangle &= \sum_j \langle j|a_j^* a_j|j\rangle \\ &= \sum_j |a_j|^2 \\ &= 1 \quad . \end{aligned} \tag{16.11}$$

The *expectation value* of an operator is found from

$$\langle\hat{X}\rangle = \langle\psi|\hat{X}|\psi\rangle \quad , \tag{16.12}$$

with an adjoint

$$\left(\langle\psi|\hat{X}|\psi\rangle\right)^{\dagger} = \langle\psi|\hat{X}^{\dagger}|\psi\rangle \quad . \tag{16.13}$$

Taking the expectation value of an eigenfunction

$$\begin{aligned} \langle n|\hat{X}|n\rangle &= \langle n|x_n|n\rangle \\ &= x_n\langle n|n\rangle \\ &= x_n \quad , \end{aligned} \tag{16.14}$$

and its adjoint

$$\begin{aligned}\left(\langle n|\hat{X}|n\rangle\right)^{\dagger} &= \langle n|\hat{X}^{\dagger}|n\rangle \\ &= \langle n|x_n^*|n\rangle \\ &= x_n^*\langle n|n\rangle \\ &= x_n^* \quad , \end{aligned} \tag{16.15}$$

shows that if the observable is real valued ($x_n = x_n^*$) then $\hat{X}^{\dagger} = \hat{X}$. An operator with this property is said to be *Hermitian*. Since the real world has real-valued observables, measurable quantities are associated with Hermitian operators.

The *outer product* of two state vectors is an operator

$$\begin{aligned}|\psi\rangle\langle\varphi| &= \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \end{pmatrix} \times (\varphi_1^* \ \varphi_2^* \ \cdots) \\ &= \begin{pmatrix} \psi_1\varphi_1^* & \psi_1\varphi_2^* & \cdots \\ \psi_2\varphi_1^* & \psi_2\varphi_2^* & \cdots \\ \vdots & \vdots & \end{pmatrix} \quad . \end{aligned} \tag{16.16}$$

The outer product of two eigenvectors

$$\hat{P}_n = |n\rangle\langle n| \tag{16.17}$$

is called a *projector* because of its action on a wave function

$$\begin{aligned}\hat{P}_n|\psi\rangle &= |n\rangle\langle n| \sum_m a_m|m\rangle \\ &= \sum_m a_m|n\rangle\langle n|m\rangle \\ &= a_n|n\rangle\langle n|n\rangle \\ &= a_n|n\rangle \quad . \end{aligned} \tag{16.18}$$

The expectation value of a projector is just the likelihood that a measurement will result in that state

$$\begin{aligned}\langle\psi|\hat{P}_n|\psi\rangle &= \sum_m \langle m|a_m^*|n\rangle\langle n|a_m|m\rangle \\ &= \sum_m |a_m|^2|\langle n|m\rangle|^2 \\ &= |a_n|^2 \quad . \end{aligned} \tag{16.19}$$

Projectors can be used to write an operator in a *spectral representation* in terms of its eigenvalues

$$\hat{X} = \sum_n x_n|n\rangle\langle n| \quad . \tag{16.20}$$

This provides a natural way to define a function of an operator as

$$f(\hat{X}) = \sum_n f(x_n)|n\rangle\langle n| \quad . \tag{16.21}$$

The *commutator* of two operators is defined to be

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} \quad . \tag{16.22}$$

If the commutator vanishes then the eigenvectors of $\hat{A}$ and $\hat{B}$ will be shared,

$$\begin{aligned} \hat{A}\hat{B} - \hat{B}\hat{A} = 0 \;\Rightarrow\; \hat{A}\hat{B}|ab\rangle &= \hat{B}\hat{A}|ab\rangle \\ \hat{A}b|ab\rangle &= \hat{B}a|ab\rangle \\ ab|ab\rangle &= ba|ab\rangle \quad , \end{aligned} \tag{16.23}$$

but in general this need not be the case.

Consider two operators that do not commute, $[\hat{A}, \hat{B}] = i\hat{C}$, where $\hat{C}$ is written with a prefactor of i for convenience in the following calculation. If we define new operators by subtracting off the expectation value $\delta\hat{A} \equiv \hat{A} - \langle\hat{A}\rangle,\ \delta\hat{B} \equiv \hat{B} - \langle\hat{B}\rangle$, then these operators will satisfy the same commutation relation $[\delta\hat{A}, \delta\hat{B}] = i\hat{C}$ because the scalar expectation values of the operators do commute. The expected value of the commutator is then

$$\begin{aligned} \langle\delta\hat{A}\ \delta\hat{B}\rangle - \langle\delta\hat{B}\ \delta\hat{A}\rangle &= i\langle\hat{C}\rangle \\ \left[\langle\delta\hat{A}\ \delta\hat{B}\rangle - \langle\delta\hat{B}\ \delta\hat{A}\rangle\right]^2 &= -\langle\hat{C}\rangle^2 \\ \langle\delta\hat{A}\ \delta\hat{B}\rangle^2 + \langle\delta\hat{B}\ \delta\hat{A}\rangle^2 & \\ -2\langle\delta\hat{A}\ \delta\hat{B}\rangle\langle\delta\hat{B}\ \delta\hat{A}\rangle &= -\langle\hat{C}\rangle^2 \\ \langle\delta\hat{A}\ \delta\hat{B}\rangle\langle\delta\hat{B}\ \delta\hat{A}\rangle &= \frac{1}{2}\left[\langle\delta\hat{A}\ \delta\hat{B}\rangle^2 + \langle\delta\hat{B}\ \delta\hat{A}\rangle^2 + \langle\hat{C}\rangle^2\right] \\ &= \frac{1}{2}\left[\langle\delta\hat{A}\ \delta\hat{B} + \delta\hat{B}\ \delta\hat{A}\rangle^2 - 2\langle\delta\hat{A}\ \delta\hat{B}\rangle\langle\delta\hat{B}\ \delta\hat{A}\rangle + \langle\hat{C}\rangle^2\right] \\ &= \frac{1}{4}\left[\langle\delta\hat{A}\ \delta\hat{B} + \delta\hat{B}\ \delta\hat{A}\rangle^2 + \langle\hat{C}\rangle^2\right] \quad . \end{aligned} \tag{16.24}$$

Using the Cauchy–Schwarz inequality yet again, the expectation of a product will be bounded by the product of the expectations

$$\langle\delta\hat{A}\ \delta\hat{B}\rangle\langle\delta\hat{B}\ \delta\hat{A}\rangle \leq \langle\delta\hat{A}^2\rangle\langle\delta\hat{B}^2\rangle \quad . \tag{16.25}$$

Therefore

$$\langle\delta\hat{A}^2\rangle\langle\delta\hat{B}^2\rangle \geq \frac{1}{4}\left[\langle\delta\hat{A}\ \delta\hat{B} + \delta\hat{B}\ \delta\hat{A}\rangle^2 + \langle\hat{C}\rangle^2\right] \quad . \tag{16.26}$$

If the anticommutator on the right hand side vanishes it sets a lower bound for the inequality of

$$\langle\delta\hat{A}^2\rangle\langle\delta\hat{B}^2\rangle \geq \frac{1}{4}\langle\hat{C}\rangle^2 \tag{16.27}$$

or

$$\langle\delta\hat{A}^2\rangle^{1/2}\langle\delta\hat{B}^2\rangle^{1/2} \equiv \Delta A \Delta B \geq \frac{1}{2}|\langle\hat{C}\rangle| \quad . \tag{16.28}$$

This is the *Heisenberg uncertainty principle*. The left hand side is the product of the expected spread around the expectation value of the operators, and the right hand side is the expectation value of the commutator. If two operators do not commute then

determining one of them more precisely imposes less certainty on the distribution of outcomes for the other. Because position and momentum don't commute (equation 16.48)

$$\Delta x \Delta p \geq \frac{\hbar}{2} \quad . \tag{16.29}$$

If we know exactly where a particle is then we have no idea how fast it is moving, or *vice versa*.

The most important operator of all is the *Hamiltonian*, the sum of the operators for the potential $\hat{T}$ and kinetic energies $\hat{V}$

$$\hat{\mathcal{H}} = \hat{T} + \hat{V} \quad . \tag{16.30}$$

In terms of it, *Schrödinger's equation* gives the time evolution of a wave function,

$$i\hbar \frac{d}{dt}|\psi(t)\rangle = \hat{\mathcal{H}}|\psi(t)\rangle \quad . \tag{16.31}$$

Because $\hat{\mathcal{H}}$ is Hermitian, the adjoint of the Hamiltonian is

$$\begin{aligned} -i\hbar \frac{d}{dt}\langle\psi(t)| &= \langle\psi(t)|\hat{\mathcal{H}}^{\dagger} \\ &= \langle\psi(t)|\hat{\mathcal{H}} \quad . \end{aligned} \tag{16.32}$$

The time derivative of the expectation value of an operator that does not depend on time can therefore be found from the chain rule,

$$\begin{aligned} i\hbar \frac{d}{dt}\langle\hat{A}\rangle &= i\hbar \frac{d}{dt}\langle\psi(t)|\hat{A}|\psi(t)\rangle \\ &= i\hbar\langle\psi(t)|\hat{A}\frac{d|\psi(t)\rangle}{dt} + i\hbar\frac{d\langle\psi(t)|}{dt}\hat{A}|\psi(t)\rangle \\ &= \langle\psi(t)|\hat{A}\hat{\mathcal{H}}|\psi(t)\rangle - \langle\psi(t)|\hat{\mathcal{H}}\hat{A}|\psi(t)\rangle \\ &= \langle\psi(t)|[\hat{A}, \hat{\mathcal{H}}]|\psi(t)\rangle \\ &= \langle[\hat{A}, \hat{\mathcal{H}}]\rangle \quad . \end{aligned} \tag{16.33}$$

This is *Ehrenfest's Theorem*; Problem 16.1 extends it to time-dependent operators. One consequence is that if an operator commutes with the Hamiltonian its observable is a constant of the motion. Another consequence follows from combining equations (16.28) and (16.33) to find

$$\Delta\hat{A}\Delta\hat{\mathcal{H}} \geq \frac{\hbar}{2}\left|\frac{d\langle\hat{A}\rangle}{dt}\right| \quad . \tag{16.34}$$

In terms of this derivative, a time Δt can be defined as

$$\left|\frac{d\langle\hat{A}\rangle}{dt}\right| \equiv \frac{\Delta\hat{A}}{\Delta t} = \frac{\langle\delta\hat{A}^2\rangle^{1/2}}{\Delta t} \quad . \tag{16.35}$$

Δt is the time it takes the expected value of an operator, at its instantaneous rate of change, to differ by its standard deviation. Recognizing that $\Delta\hat{\mathcal{H}} \equiv \Delta E$ is the expected energy spread,

$$\Delta\hat{A}\Delta E \geq \frac{\hbar}{2}\frac{\Delta\hat{A}}{\Delta t} \quad , \tag{16.36}$$

and cancelling $\Delta\hat{A}$ gives

$$\Delta E \Delta t \geq \frac{\hbar}{2} \quad . \tag{16.37}$$

This is the *energy–time uncertainty relation*. It fundamentally differs from the other uncertainty relationships because time appears in quantum mechanics as a parameter rather than an operator. Equation (16.37) can be interpreted as putting a quantum limit on how the energy distribution in a system bounds the rate at which an observable property of the system can change, although there are *many* other ways to (mis)define and (mis)interpret it [Aharonov & Bohm, 1961; Peres, 1993].

The *evolution operator* $\hat{U}(t)$ is the one that advances a wave function in time:

$$|\psi(t)\rangle = \hat{U}(t)|\psi(0)\rangle \quad . \tag{16.38}$$

Differentiating both sides and comparing to equation (16.31) shows that $\hat{U}$ satisfies

$$\begin{aligned} i\hbar\frac{d}{dt}|\psi(t)\rangle &= i\hbar\frac{d}{dt}\hat{U}(t)|\psi(0)\rangle \\ \hat{\mathcal{H}}|\psi(t)\rangle &= i\hbar\frac{d}{dt}\hat{U}(t)|\psi(0)\rangle \\ \hat{\mathcal{H}}\hat{U}(t)|\psi(0)\rangle &= i\hbar\frac{d}{dt}\hat{U}(t)|\psi(0)\rangle \quad . \end{aligned} \tag{16.39}$$

The evolution operator therefore is a solution to

$$i\hbar\frac{d}{dt}\hat{U}(t) = \hat{\mathcal{H}}\hat{U}(t) \quad . \tag{16.40}$$

If the Hamiltonian does not explicitly depend on time this can be integrated to find that

$$\hat{U}(t) = e^{-i\hat{\mathcal{H}}t/\hbar} \quad , \tag{16.41}$$

where the exponential of an operator can be defined by equation (16.21) or a power series expansion

$$e^{-i\hat{\mathcal{H}}t/\hbar} = \hat{1} - \frac{it}{\hbar}\hat{\mathcal{H}} + \frac{1}{2!}\left(\frac{it}{\hbar}\right)^2 \hat{\mathcal{H}}\hat{\mathcal{H}} + \cdots \tag{16.42}$$

(where $\hat{1}$ is the identity operator).

We know that wave functions always start out normalized:

$$\langle\psi(0)|\psi(0)\rangle = 1 \quad . \tag{16.43}$$

At a future time the inner product is

$$\langle\psi(t)|\psi(t)\rangle = \langle\psi(0)|\hat{U}(t)^\dagger\hat{U}(t)|\psi(0)\rangle \quad , \tag{16.44}$$

which must still equal 1. Therefore

$$\hat{U}^\dagger\hat{U} = \hat{I} \quad , \tag{16.45}$$

where $\hat{I}$ is the identity operator (matrix). An operator with this property is said to be *unitary*, and corresponds to a rotation in Hilbert space.

16.1.2 Angular Momentum

The non-obvious properties of quantum mechanical angular momentum are essential to understanding quantum information. They are found by generalizing the classical definition. The quantum position operator is simple in the position basis, just giving the coordinates

$$\hat{\vec{x}} = \vec{x} \quad . \tag{16.46}$$

The momentum operator is less trivial in the position basis; it is equal to the spatial gradient of the wave function

$$\hat{\vec{p}} = -i\hbar\nabla \tag{16.47}$$

(while this can be motivated, it's simplest to take it as an experimental fact). Therefore position and momentum do not commute:

$$\begin{aligned}
[\hat{\vec{x}}, \hat{\vec{p}}]|\psi\rangle &= -\vec{x}i\hbar\nabla|\psi\rangle + i\hbar\nabla\vec{x}|\psi\rangle \\
&= -\vec{x}i\hbar\nabla|\psi\rangle + i\hbar|\psi\rangle + i\hbar\vec{x}\nabla|\psi\rangle \\
&= i\hbar|\psi\rangle \\
[\hat{\vec{x}}, \hat{\vec{p}}] &= i\hbar \quad .
\end{aligned} \tag{16.48}$$

The angular momentum operator is found from

$$\hat{\vec{L}} = \hat{\vec{x}} \times \hat{\vec{p}} \quad . \tag{16.49}$$

Writing out the components shows that

$$[\hat{L}_l, \hat{x}_m] = i\hbar\epsilon_{lmn}\hat{x}_n \quad , \tag{16.50}$$

$$[\hat{L}_l, \hat{p}_m] = i\hbar\epsilon_{lmn}\hat{p}_n \quad , \tag{16.51}$$

and

$$[\hat{L}_l, \hat{L}_m] = i\hbar\epsilon_{lmn}\hat{L}_n \quad . \tag{16.52}$$

The last commutator is satisfied by any quantity that behaves like an angular momentum, including the spin degrees of freedom.

Because the components of angular momentum do not commute it is not possible for a wave function to simultaneously be in an eigenstate of all of them. By convention, the direction that is an angular momentum eigenstate is taken to be z. But evaluating the components does show that the angular momentum operator commutes with the total angular momentum $\hat{L}^2 = \hat{\vec{L}} \cdot \hat{\vec{L}}$

$$[\hat{L}^2, \hat{\vec{L}}] = 0 \quad . \tag{16.53}$$

Therefore a wave function can simultaneously be an eigenfunction of the total angular momentum and its component in one direction. If l is the quantum number for the total angular momentum and m the quantum number for the z component, an angular momentum eigenstate can be written as $|l, m\rangle$. For the coming calculations it will be convenient to define these indices by

$$\begin{aligned}
\hat{L}^2|l, m\rangle &= \hbar^2 l(l+1)|l, m\rangle \\
\hat{L}_z|l, m\rangle &= \hbar m|l, m\rangle \quad .
\end{aligned} \tag{16.54}$$

Two useful new operators $\hat{L}_+$ and $\hat{L}_-$ can be defined in terms of the orthogonal directions

$$\hat{L}_\pm = \hat{L}_x \pm i\hat{L}_y \quad . \tag{16.55}$$

From this definition its straightforward to show that

$$\hat{L}_-^\dagger = \hat{L}_+ \quad , \tag{16.56}$$

$$\hat{L}_+\hat{L}_- = \hat{L}^2 - \hat{L}_z^2 + \hbar\hat{L}_z \quad , \tag{16.57}$$

$$[\hat{L}_z, \hat{L}_\pm] = \pm\hbar\hat{L}_\pm \quad , \tag{16.58}$$

and

$$[\hat{L}_+, \hat{L}_-] = 2\hbar\hat{L}_z \quad . \tag{16.59}$$

Because of the adjoint relationship, the product of these operators must be positive

$$\begin{aligned} \langle l,m|\hat{L}_+\hat{L}_-|l,m\rangle &= \langle l,m|\hat{L}_-^\dagger\hat{L}_-|l,m\rangle \\ &= |\hat{L}_-|l,m\rangle|^2 \\ &\geq 0 \quad . \end{aligned} \tag{16.60}$$

But we also know that

$$\begin{aligned} \langle l,m|\hat{L}_+\hat{L}_-|l,m\rangle &= \langle l,m|\hat{L}^2 - \hat{L}_z^2 + \hbar\hat{L}_z|l,m\rangle \\ &= \hbar^2[l(l+1) - m^2 + m]\ \langle l,m|l,m\rangle \\ &= \hbar^2[l(l+1) - m^2 + m] \quad . \end{aligned} \tag{16.61}$$

Therefore

$$l(l+1) - m^2 + m \geq 0 \quad . \tag{16.62}$$

And taking the operators in the opposite order gives

$$l(l+1) - m^2 - m \geq 0 \quad . \tag{16.63}$$

That means that

$$l^2 + l \geq m^2 \pm m \quad , \tag{16.64}$$

which can only be satisfied if

$$-l \leq m \leq l \quad . \tag{16.65}$$

Now consider the following sequence of operators:

$$\begin{aligned} \hat{L}_z\hat{L}_-|l,m\rangle &= (\hat{L}_-\hat{L}_z - \hbar\hat{L}_-)|l,m\rangle \\ &= \hbar(m-1)\hat{L}_-|l,m\rangle \quad . \end{aligned} \tag{16.66}$$

$\hat{L}_-$ gives a new eigenfunction of $\hat{L}_z$ with m reduced by 1; for this reason it is called a *lowering operator*. The normalization can be found from equation (16.61) to be

$$\hat{L}_-|l,m\rangle = \hbar\sqrt{l(l+1) - m(m-1)}|l,m-1\rangle \quad . \tag{16.67}$$

Likewise, $\hat{L}_+$ is a *raising operator*

$$\hat{L}_+|l,m\rangle = \hbar\sqrt{l(l+1) - m(m+1)}|l,m+1\rangle \quad . \tag{16.68}$$

Since $-l \leq m \leq l$, and we've just found that m can be changed by integer increments n,

$$\begin{aligned} -l + nm &= l \\ \frac{nm}{2} &= l \quad . \end{aligned} \tag{16.69}$$

l must be either an integer $(0, 1, \ldots)$ or a half-integer $(1/2, 3/2, \ldots)$.

A spin-1/2 particle like an electron or a proton has $l = 1/2$, and two possible m states which can be written equivalently as

$$\begin{aligned} |l, m\rangle = |1/2, 1/2\rangle = |\uparrow\rangle &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ |l, m\rangle = |1/2, -1/2\rangle = |\downarrow\rangle &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad . \end{aligned} \tag{16.70}$$

In the last basis the angular momentum operator can be written in terms of the *Pauli spin matrices*

$$\hat{\vec{L}} = \frac{\hbar}{2}\hat{\vec{\sigma}}$$

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \hat{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad , \tag{16.71}$$

which can be verified by checking their commutation relations.

The spin magnetic moment $\vec{\mu}$ of a particle is related to its angular momentum $\vec{L}$ by the gyromagnetic ratio $\vec{\mu} = \gamma\vec{L}$ (Section 10.4). The moment can thus be determined by the operator

$$\begin{aligned} \hat{\vec{\mu}} &= \gamma\hat{\vec{L}} \\ &= \gamma\frac{\hbar}{2}\hat{\vec{\sigma}} \quad . \end{aligned} \tag{16.72}$$

If the spin density is n, the macroscopic spin magnetization is therefore

$$\vec{M} = n\gamma\frac{\hbar}{2}\langle\psi|\hat{\vec{\sigma}}|\psi\rangle \quad . \tag{16.73}$$

Now consider two systems labelled a and b, with angular momentum operators $\hat{\vec{L}}(a)$ and $\hat{\vec{L}}(b)$. Because they satisfy the angular momentum relations individually, their sum

$$\hat{\vec{L}} = \hat{\vec{L}}(a) + \hat{\vec{L}}(b) \tag{16.74}$$

will also. But because

$$\begin{aligned} \hat{L}^2 &= \left[\hat{\vec{L}}(a) + \hat{\vec{L}}(b)\right]^2 \\ &= \hat{L}^2(a) + \hat{L}^2(b) + 2\hat{\vec{L}}(a) \cdot \hat{\vec{L}}(b) \\ &= \hat{L}^2(a) + \hat{L}^2(b) + 2\hat{L}_z(a)\hat{L}_z(b) + \hat{L}_+(a)\hat{L}_-(b) + \hat{L}_-(a)\hat{L}_+(b) \end{aligned} \tag{16.75}$$

the eigenfunctions for systems a and b will in general not be angular momentum eigenfunctions of the combined system. For spin 1/2 there are four eigenfunctions of the

individual systems

$$|\uparrow\uparrow\rangle\ ,\ |\uparrow\downarrow\rangle\ ,\ |\downarrow\uparrow\rangle\ ,\ |\downarrow\downarrow\rangle\ \ , \tag{16.76}$$

and the magnitude of the total angular momentum is from equation (16.75)

$$\hat{L}^2 = \frac{3}{4} + 2\hat{L}_z(a)\hat{L}_z(b) + \hat{L}_+(a)\hat{L}_-(b) + \hat{L}_-(a)\hat{L}_+(b) \quad . \tag{16.77}$$

Substitution shows that the eigenfunctions of the combined systems are

$$\begin{aligned}
|l=0, m=0\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) \\
|l=1, m=-1\rangle &= |\downarrow\downarrow\rangle \\
|l=1, m=0\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) \\
|l=1, m=1\rangle &= |\uparrow\uparrow\rangle \quad .
\end{aligned} \tag{16.78}$$

The $l = 0$ state is called a *singlet*, and the other three are called *triplet* states. These are the symmetric and antisymmetric spin states that we used in Chapter 13 to explain magnetism. For the more general case of adding arbitrary angular momenta it is necessary to use *Clebsch–Gordon coefficients* to find the expansion of the total angular momentum eigenfunctions in terms of the individual basis functions.

The angular momentum operator, not surprisingly, has an intimate connection with rotations. Consider a *rotation operator* $\hat{R}(\vec{\theta})$ that returns the wave function rotated by an angle θ around the axis of $\vec{\theta}$. Rotating the wave function forward is the same as rotating its argument backwards, and in the limit of a small angle the coordinate rotation matrix can be approximated by a cross product

$$\begin{aligned}
\hat{R}(\delta\vec{\theta})|\psi(\vec{x})\rangle &\approx |\psi(\vec{x} - \delta\vec{\theta}\times\vec{x})\rangle \\
&\approx |\psi(\vec{x})\rangle - (\delta\vec{\theta}\times\vec{x})\cdot\nabla|\psi(\vec{x})\rangle \\
&= |\psi(\vec{x})\rangle - \delta\vec{\theta}\cdot\vec{x}\times\nabla|\psi(\vec{x})\rangle \\
&= |\psi(\vec{x})\rangle - \delta\vec{\theta}\cdot\vec{x}\times\frac{1}{-i\hbar}\hat{\vec{p}}|\psi(\vec{x})\rangle \\
&= |\psi(\vec{x})\rangle - \frac{i}{\hbar}\delta\vec{\theta}\cdot\hat{\vec{L}}|\psi(\vec{x})\rangle \\
\hat{R}(\delta\vec{\theta}) &= \hat{I} - \frac{i}{\hbar}\delta\vec{\theta}\cdot\hat{\vec{L}} \quad .
\end{aligned} \tag{16.79}$$

The angular momentum operator has appeared as the generator of an infinitesimal rotation. A finite rotation can be built up from the product of many inifinitesimal rotations, so that if we write $\vec{\theta} = N\ \delta\vec{\theta}$ and use the limit

$$\lim_{N\to\infty}\left[1 + \frac{x}{N}\right]^N = e^x \tag{16.80}$$

we find the rotation operator to be

$$\hat{R}(\vec{\theta}) = e^{-i\vec{\theta}\cdot\vec{L}/\hbar} \quad . \tag{16.81}$$

Plugging in the Pauli spin matrices gives the rotation operators for each axis

(Problem 16.4):

$$\hat{R}_x(\theta) = \begin{pmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$$
$$\hat{R}_y(\theta) = \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$$
$$\hat{R}_z(\theta) = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix} \quad . \tag{16.82}$$

These are going to be essential ingredients in implementing quantum logic.

16.1.3 Density Matrices

Uncertainty is central to quantum mechanics through the unpredictability of the outcome of a measurement. It is also central to thermodynamics, in the probability distribution over allowed states. These have very different characters, though: the former is a fundamental limit while the latter reflects incomplete knowledge about a system with many degrees of freedom. These can be combined by using the *density matrix* $\hat{\rho}$.

A *pure state* is one for which we know the exact state $|n\rangle$; its density matrix is just defined to be the projector

$$\hat{\rho} = |n\rangle\langle n| \quad . \tag{16.83}$$

A *mixed state* is one that has a classical probability distribution over possible quantum states. If the classical probability for a mixed state to be in the quantum state $|\psi_i\rangle$ is p_i then the density matrix is a weighted sum over the projection operators

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i| \quad . \tag{16.84}$$

The *trace* of an operator is defined by

$$\mathrm{Tr}(\hat{X}) = \sum_n \langle n|\hat{X}|n\rangle \quad . \tag{16.85}$$

It has the property of *cyclic invariance*

$$\mathrm{Tr}(\hat{X}\hat{Y}) = \mathrm{Tr}(\hat{Y}\hat{X}) \quad , \tag{16.86}$$

and is unchanged under a change of basis. In a matrix representation of an operator the trace is equal to the sum of the diagonal elements. Evaluating the trace of a density matrix in the basis in which it is diagonal shows that it is normalized:

$$\begin{aligned} \mathrm{Tr}(\hat{\rho}) &= \sum_n \rho_{nn} \\ &= \sum_n \langle n|\hat{\rho}|n\rangle \\ &= \sum_n \langle n| \sum_{n'} p_{n'} |n'\rangle\langle n'|n\rangle \end{aligned}$$

$$= \sum_n p_n$$
$$= 1 \quad . \tag{16.87}$$

But $\mathrm{Tr}(\hat{\rho}^2) \leq 1$, with equality holding for pure states (Problem 16.2). This provides a way to distinguish between the density matrices for pure and mixed states.

In terms of the trace over $\hat{\rho}$, the expectation value of an observable is

$$\begin{aligned}\langle \hat{X} \rangle &= \mathrm{Tr}(\hat{\rho}\hat{X}) \\ &= \sum_n \langle n| \sum_i p_i |\psi_i\rangle\langle\psi_i|\hat{X}|n\rangle \\ &= \sum_n \langle n| \sum_i p_i \sum_{n'} a_{in'}|n'\rangle\langle n'|a^*_{in'}\hat{X}|n\rangle \\ &= \sum_n \sum_i p_i a_{in}\langle n|a^*_{in'}\hat{X}|n\rangle \\ &= \sum_i p_i \langle\psi_i|\hat{X}|\psi_i\rangle \quad .\end{aligned} \tag{16.88}$$

This is a classically weighted sum over the quantum expectation.

The time derivative of the density matrix

$$\begin{aligned}\frac{d\hat{\rho}}{dt} &= \sum_i p_i \left[\frac{d|\psi_i\rangle}{dt}\langle\psi_i| + |\psi_i\rangle\frac{d\langle\psi_i|}{dt}\right] \\ &= \sum_i p_i \left[\frac{1}{i\hbar}\hat{\mathcal{H}}|\psi_i\rangle\langle\psi_i| + |\psi_i\rangle\frac{1}{-i\hbar}\langle\psi_i|\hat{\mathcal{H}}\right] \\ &= \frac{1}{i\hbar}[\hat{\mathcal{H}}, \hat{\rho}]\end{aligned} \tag{16.89}$$

is given by its commutator with the Hamiltonian. This is the *Liouville–von Neumann* evolution equation. The time dependence of the density matrix can also be found from equation (16.41) to be the unitary transformation

$$\begin{aligned}\hat{\rho}(t) &= \sum_i p_i |\psi_i(t)\rangle\langle\psi_i(t)| \\ &= \sum_i p_i e^{-i\hat{\mathcal{H}}t/\hbar}|\psi_i(0)\rangle\langle\psi_i(0)|e^{i\hat{\mathcal{H}}t/\hbar} \\ &= e^{-i\hat{\mathcal{H}}t/\hbar}\hat{\rho}(0)e^{i\hat{\mathcal{H}}t/\hbar} \\ &= \hat{U}(t)\hat{\rho}(t)\hat{U}^\dagger(t) \quad .\end{aligned} \tag{16.90}$$

Unitary evolution is a serious constraint on quantum dynamics. It cannot change the trace, eigenvalue spectrum, or entropy of a density matrix. This means that information cannot be created or erased, just rearranged. How can this be reconciled with our everyday experience of these things? The answer is that while the universe as a whole is believed to be unitary, subsystems need not be.

If there are two quantum systems with state vectors $|u\rangle$ and $|v\rangle$, then a state vector in the combined system is given by the *tensor product* $|u\rangle \otimes |v\rangle$, which can be abbreviated as $|u\rangle|v\rangle$ or just $|uv\rangle$. Operators distribute over the tensor product

$$\left(\hat{U} \otimes \hat{V}\right)\left(|u\rangle \otimes |v\rangle\right) = \hat{U}|u\rangle \otimes \hat{V}|v\rangle \quad ; \tag{16.91}$$

in a matrix representation the tensor product of two operators is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \otimes \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} a\alpha & a\beta & b\alpha & b\beta \\ a\gamma & a\delta & b\gamma & b\delta \\ c\alpha & c\beta & d\alpha & d\beta \\ c\gamma & c\delta & d\gamma & d\delta \end{pmatrix} \quad . \tag{16.92}$$

Taking a *partial trace* over a subsystem returns the density matrix for the other part:

$$\begin{aligned} \mathrm{Tr}_v(\hat{\rho}_{uv}) &= \mathrm{Tr}_v \left(\sum_{ij} p_{ij} |u_i v_j\rangle\langle u_i v_j| \right) \\ &= \sum_n \langle v_n | \sum_{ij} p_{ij} |u_i v_j\rangle\langle u_i v_j | | v_n \rangle \\ &= \sum_{ij} p_{ij} |u_i\rangle\langle u_i| \\ &= \sum_i p_i |u_i\rangle\langle u_i| \\ &= \hat{\rho}_u \quad . \end{aligned} \tag{16.93}$$

The expectation value of an observable in a subsystem is thus found by tracing out the rest of the system

$$\begin{aligned} \mathrm{Tr}_u(\hat{U}\hat{\rho}_u) &= \mathrm{Tr}_u \left(\hat{U} \mathrm{Tr}_v(\hat{\rho}_{uv}) \right) \\ &= \mathrm{Tr}_{uv} \left((\hat{U} \otimes \hat{I}_v) \hat{\rho}_{uv} \right) \quad . \end{aligned} \tag{16.94}$$

While unitary evolution cannot create or destroy information, it can move it between subsystems. If a partial trace is taken over just a subsystem of interest, its dynamics no longer must be unitary. One system could be a computer, and the other the rest of the universe. When the computer appears to erase a bit it really just moves the information to unobserved degrees of freedom, so that the combined system remains unitary.

Through this mechanism the operators, like projectors, that describe measurements need not be unitary because information can be exchanged between the system being measured and the measurement apparatus. Let $\hat{M}_n$ be the measurement operator corresponding to obtaining the value n, for example a projector $|\uparrow\rangle\langle\uparrow|$ onto the up-state of a spin. The probability of this outcome is found from the magnitude

$$\begin{aligned} p(n) &= \left| \hat{M}_n |\psi\rangle \right| \\ &= \langle \psi | \hat{M}_n^\dagger \hat{M}_n | \psi \rangle \quad . \end{aligned} \tag{16.95}$$

If this does occur, the new state vector is found by normalizing the result of this operator

$$\begin{aligned} |\psi\rangle &\xrightarrow{n} \frac{\hat{M}_n |\psi\rangle}{\left| \hat{M}_n |\psi\rangle \right|} \\ &= \frac{\hat{M}_n |\psi\rangle}{\left[\langle \psi | \hat{M}_n^\dagger \hat{M}_n | \psi \rangle \right]^{1/2}} \quad . \end{aligned} \tag{16.96}$$

Likewise, the density matrix becomes

$$\hat{\rho} \xrightarrow{n} \frac{\hat{M}_n \hat{\rho} \hat{M}_n^\dagger}{\mathrm{Tr}\left(\hat{M}_n \hat{\rho} \hat{M}_n^\dagger\right)} \quad . \tag{16.97}$$

In thermal equilibrium the state probabilities are just given by Boltzmann factors, so a thermalized density matrix written in terms of energy eigenstates $|n\rangle$ with eigenvalues E_n is

$$\begin{aligned} \hat{\rho}_{\text{thermal}} &= \sum_n p_n |n\rangle\langle n| \\ &= \sum_n \frac{e^{-\beta E_n}}{\sum_m e^{-\beta E_m}} |n\rangle\langle n| \\ &= \frac{e^{-\beta \hat{\mathcal{H}}}}{\mathrm{Tr}(e^{-\beta \hat{\mathcal{H}}})} \\ &= \frac{e^{-\beta \hat{\mathcal{H}}}}{\mathcal{Z}} \quad , \end{aligned} \tag{16.98}$$

where $\beta = 1/kT$ and $\mathcal{Z}$ is the partition function. In this basis the density matrix is diagonal. The possibility of processing information quantum mechanically rests on the availability of all of the off-diagonal terms. These represent quantum *coherence*, which eventually *decoheres* by interaction with the environment [Zurek, 1998]. This is the essential challenge in working with quantum information: while it's necessary to manipulate it to do anything useful, it must carefully be isolated to protect it from decoherence. You will break a quantum computer if you look at it too hard, which given the fragility of quantum coherence is not very hard at all.

16.2 INFORMATION

The quantum theory of information in many ways looks like its classical counterpart, with the classical entropy sum

$$H(p) = -\sum_n p_n \log p_n \tag{16.99}$$

replaced by a trace over the density matrix

$$S(\hat{\rho}) = -\mathrm{Tr}(\hat{\rho} \log \hat{\rho}) \quad . \tag{16.100}$$

Using this *von Neumann entropy*, corresponding noiseless and noisy coding theorems hold [Jozsa & Schumacher, 1994; Lloyd, 1997].

Don't be lulled into a false sense of security, however: quantum information behaves profoundly unlike its classical counterpart. A bit of quantum of information is called a *qubit*. It might reside in two orthogonal polarization states of a linearly polarized photon, or the spin-up and spin-down orientations of a proton in a magnetic field. Such qubit eigenstates can be written as $|0\rangle$ and $|1\rangle$, where 0 and 1 index the two orthogonal basis states. Unlike a classical bit which must be a 0 or 1, or an analog degree of freedom that

could be between 0 and 1, a qubit can be in a *superposition*

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{16.101}$$

in which it simultaneously has $|0\rangle$ and $|1\rangle$ components.

The configuration of a two-bit classical register is specified by giving the value of the bits (which can be 00, 01, 10, or 11); describing the state of a two-qubit register requires $2^2 = 4$ coefficients $\alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle$. An N-bit classical register can store N bits, but the configuration of N qubits is specified by 2^N coefficients ($2^N - 1$ actually, because of the normalization constraint). This ability to be in an arbitrary superposition lets a qubit represent exponentially more information than a classical bit. It introduces a quantum notion of parallelism because a quantum computer can operate on all possible inputs at the same time.

A second difference between quantum information and classical information is that it cannot be copied. Consider an "amplifier" operator $\hat{A}$ that (if it could exist) would take an input state and produces two identical output states

$$\hat{A}|\psi\rangle = |\psi\rangle|\psi\rangle \quad . \tag{16.102}$$

If we ask it to amplify a superposition we should get two copies

$$\begin{aligned}\hat{A}\left(\alpha|0\rangle + \beta|1\rangle\right) &= \left(\alpha|0\rangle + \beta|1\rangle\right)\left(\alpha|0\rangle + \beta|1\rangle\right) \\ &= \alpha^2|0\rangle|0\rangle + \beta^2|1\rangle|1\rangle + \alpha\beta\left(|0\rangle|1\rangle + |1\rangle|0\rangle\right) \quad .\end{aligned} \tag{16.103}$$

The problem is that quantum mechanical operators must be linear. This means that our amplifier operator must also satisfy

$$\begin{aligned}\hat{A}\left(\alpha|0\rangle + \beta|1\rangle\right) &= \alpha\hat{A}|0\rangle + \beta\hat{A}|1\rangle \\ &= \alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle \quad ,\end{aligned} \tag{16.104}$$

a contradiction with equation (16.103). The only way for these equations to agree is if the state to be amplified is not a superposition in the basis used by the amplifier. But this means that its value can be determined with certainty and hence a quantum amplifier isn't even needed to generate copies. Therefore, it is not possible to copy an arbitrary quantum state. This *no-clone* theorem [Wooters & Zurek, 1982] is the foundation of quantum cryptography. It is a consequence of the unitarity of quantum mechanics, which is reversible and hence allows qubits to be rearranged but not created or destroyed (although we will see that there is a loophole that permits error correction). Fortunately, the classical study of reversible computation showed that this is indeed possible with just a small overhead in complexity [Bennett, 1988].

The most remarkable property of all is *entanglement*. This lies at the heart of the apparent mystery of quantum mechanics, establishing a spooky connection among quantum systems after they interact. If a system comprises two particles, a and b, it is said to be in a *product state* if the total wave function can be written as a product of the individual particle wave functions

$$|\psi\rangle = |a\rangle|b\rangle \quad . \tag{16.105}$$

This is not always the case. Consider a radioactive decay process the emits two spin-1/2

particles. If the system starts out with no net angular momentum it must end up with no net angular momentum, and so it must be in the singlet state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_a|\downarrow\rangle_b - |\downarrow\rangle_a|\uparrow\rangle_b\right) \quad . \tag{16.106}$$

Because this cannot be written as a product state it is said to be *entangled*. Now make a measurement of the spin orientation of particle a. Before the measurement there's an equal chance for it to be up or down; after the measurement it must be either up or down. But something unexpected has happened to b after the measurement: its state is also determined. If a turns out to be up then b is down, and *vice versa*. This holds instantaneously and independent of the distance between the particles. According to the laws of quantum mechanics b could have been sent to Mars and its state would still have been determined by a measurement of a on Earth. This disturbing effect has been confirmed experimentally with entangled photons emitted in a multiple-photon process [Aspect *et al*., 1981; Tittel *et al*., 1998]

This is called the *Einstein–Podolsky–Rosen* paradox [Einstein *et al*., 1935], and because it appears to violate the locality required by relativity it is one of the reasons why Einstein didn't like quantum mechanics. Such entangled particles are called an *EPR pair* in honor of the first people to be bothered by them. There are almost as many interpretations as there are interpreters; among the popular ones are:

- Nothing is wrong. No one disagrees with the predictions of quantum mechanics; the only problem is our intuition, which has no place here.
- Since the particles have a shared origin, they contain *hidden variables* that we don't know how to access but which define the outcome of the measurements. This would be a satisfying explanation, but a series of results starting with *Bell's Theorem* appear to rule it out [Bell, 1964; Greenberger *et al*., 1990; Peres, 1990; Mermin, 1993]. Problem 16.3 works through a distressingly simple example.
- After particle a's interaction a message can travel backwards in time to the shared origin and then forwards in time to b in the present.
- The wave function for the universe splits following a measurement into non-interacting branches, selecting out the one with the correct answer for particle b. This is Everett's *Many-Worlds* Theory [Everett, 1957].

Amidst these and the even more bizarre explanations that have been put forward [Mermin, 1985], it is important to keep in mind that

- The predictions of quantum mechanics have been experimentally verified to fabulous precision. This is a paradox about interpretation, not about what is actually calculated and observed.
- Entanglement cannot be used to send messages faster than the speed of light, much to the consternation of some research grantors and grantees. b knows the state of a following the measurement, but the state of a can't be set in advance.

Entanglement is a channel that carries purely quantum information. We will use it for teleportation in communications, and as a kind of interconnect in Hilbert space for computation.

16.3 COMMUNICATIONS

Given the tools to describe quantum information, we now turn to explore the possibilities afforded by quantum mechanical communications, from new ways to do old tasks (like cryptography) to ideas that until recently were matters of science fiction rather than fact (like teleportation).

16.3.1 Cryptography

Secure electronic transactions rely on cryptography, classical cryptography relies on number theory (Section 14.3.4), and this branch of number theory ultimately rests on very little. While there are good reasons to believe that finding prime factors is necessarily an exponentially difficult task, and hence for all practical purposes can be considered intractable, there is no proof of this. That is why there is growing appreciation of the insight that the theory of quantum measurement can be applied to protect information by physical laws rather than number theory [Wiesner, 1983].

The most important application is to the weakest link in the cryptographic chain, distributing private keys to establish a shared secret. The initial scheme of [Bennett & Brassard, 1984] is shown in Figure 16.1. Alice wants to send a key to Bob, and Eve wants to intercept it (these names are required in any discussion of cryptographic protocols).

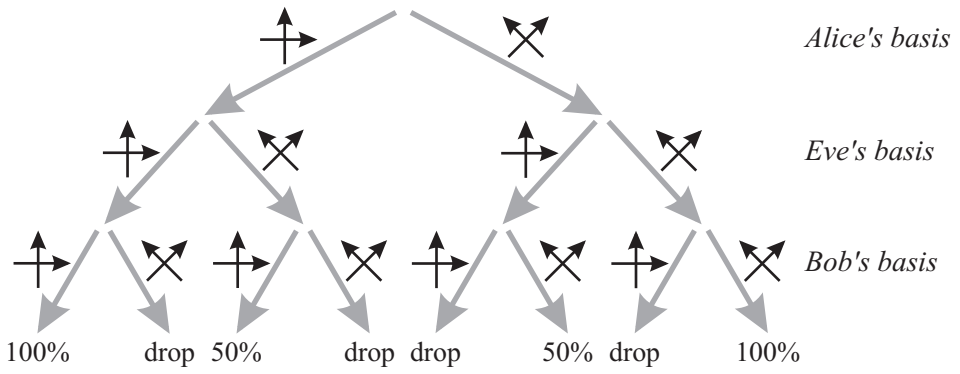

Figure 16.1. Quantum key distribution with an eavesdropper, showing the probability of Eve remaining undetected for each of the possible combinations of polarization bases.

Alice is going to communicate by sending one photon per bit to Bob, using a source that can emit single photons [Brunel *et al*., 1999]. The bit will be represented by one of two orthogonal polarization directions, and for each bit she chooses one of two sets of orientations of the axes. Bob will likewise independently choose one of the two possible orientations for his axes to measure the polarization state. After a string of bits is sent, Alice and Bob can publicly report which orientations they used. Bits for which these don't match are discarded because the result of Bob's measurements would be random. Alice and Bob then publicly compare the values of some of the bits that were generated and measured using the same axis orientation. If there is no eavesdropping then these will match.

Now consider Eve's influence. Since each bit is represented by a single photon, because

of the no-clone theorem Eve cannot surreptitiously determine the value of the bit without performing a measurement on the photon. Since she has no advance information about the random orientation of Alice's axes, she can do no better than randomly choose her own orientation, and following a measurement send a new matching photon on to Bob. If Eve is lucky enough to guess Alice's basis then she can reliably determine the polarization state and generate an identical photon to send to Bob. But if her axes don't match Alice's then Eve will obtain a random result. She has no way of telling which is which.

If Bob and Eve both match Alice's basis then Bob will receive the correct bit value and not detect Eve's presence. But if Bob's basis matches Alice's when Eve's doesn't, half of Bob's measurement on Eve's bit copies will randomly match what Alice sent, and half won't. This means that Eve has a total probability of 3/4 of being undetected on a single bit. That's pretty good odds for one bit, but if Alice and Bob compare two bits her probability of being undetected on both is $(3/4)^2$, and for N bits $(3/4)^N$. This exponential growth means that Alice and Bob need compare only a short string of values to be confident of detecting Eve's presence.

Unlike conventional cryptographic schemes that seek to obscure information, this one detects tampering with the message after it happens. While it suceeds only statistically rather than with certainty, its power comes from the exponential bound put on the probability of the eavesdropper remaining undetected. Like cryptosystems based on the (presumed) difficulty of factoring, a linear increase in effort by the sender leads to an exponential increase in the difficulty for the eavesdropper. Beyond resting on the firm foundation of the impossibility of copying quantum information, one of the appeals of quantum cryptography is the ease of understanding the scheme, although ruling out attacks by a quantum adversary has been a much more challenging task [Lo & Chau, 1999].

From the first experimental demonstration of a quantum cryptosystem [Bennett & Brassard, 1989], implementations have matured to include links over tens of kilometers using commercial telecommunications fibers [Muller *et al*., 1996], and through the atmosphere in daylight [Hughes *et al*., 2000].

16.3.2 Circuits

Quantum channels, like classical ones, require coding to take full advantage of their capabilities. In Chapter 11 we saw that a nonlinear gate plus an inverter is sufficient to generate any logical function. This can't be true quantum mechanically because the constraint of unitarity requires that the primitive gates be invertible so that it's possible to deduce their inputs from their outputs. But an analogous result does hold: any unitary transformation can be decomposed into a combination of unitary operations on a single qubit and a nonlinear two-qubit operator [Barenco *et al*., 1995].

Because a unitary transformation of a qubit preserves its norm, it can be written as a rotation about an arbitrary axis $\vec{\theta}$ by an angle θ, along with a possible overall phase factor

$$\hat{U} = e^{i\varphi}\hat{R}(\vec{\theta}) \quad . \tag{16.107}$$

The rotation can in turn be decomposed into rotations about an orthogonal set of axes by using the rotation operators in equation (16.82). This is the quantum analog of a `NOT` gate, but the continuous parameterization lets it do many more things. Problem 16.5

constructs a $\sqrt{\texttt{NOT}}$ gate that serves as an inverter when applied twice; a particularly useful single qubit gate will be the *Hadamard* transform

$$\hat{H} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad . \tag{16.108}$$

An example of a nonlinear two-input quantum gate is the CNOT (*Controlled-NOT*) transformation

$$\texttt{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{matrix} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{matrix} \quad . \tag{16.109}$$

If the first bit is down nothing happens to the second, and if the first bit is up the second one is flipped. This is a quantum generalization of XOR, made reversible by retaining the control input.

Figure 16.2 shows a simple circuit made up of a Hadamard transform followed by a CNOT. By convention, qubits A and B enter at the left and time advances along the wires towards the right, and the symbol for the CNOT indicates that A is the control qubit and B is the target. Working through the circuit, the Hadamard first puts A into a superposition. After the CNOT, if A is in the $|0\rangle$ state nothing happens to B, but if it is a $|1\rangle$ then B will be inverted. Because A is in a superposition, B is conditionally flipped based on the state of A. The value of B is not determined until that of A is (or *vice versa*). This circuit creates an entangled pair.

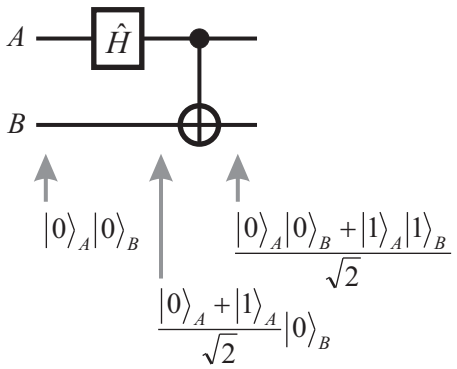

Figure 16.2. Hadamard and CNOT gates in a quantum circuit.

16.3.3 Teleportation

Quantum channels can send quantum as well as classical information, with remarkable consequences. Consider the circuit shown in Figure 16.3. It starts with Hadamard and CNOT gates to create an EPR pair from Alice's and Bob's qubits. Then, Alice stays on the Earth while Bob flys to the Moon (for example). Alice now wishes to send a new qubit

to Bob. She can't measure it and communicate its state classically because that would force it into the measurement basis. But she can use a second CNOT gate to entangle the qubit with her half of the EPR pair, which is in turn entangled with Bob's qubit. She does this in the opposite order of the EPR circuit, following the CNOT with a Hadamard transform. Finally, Alice does perform a measurement to determine the state of her two qubits.

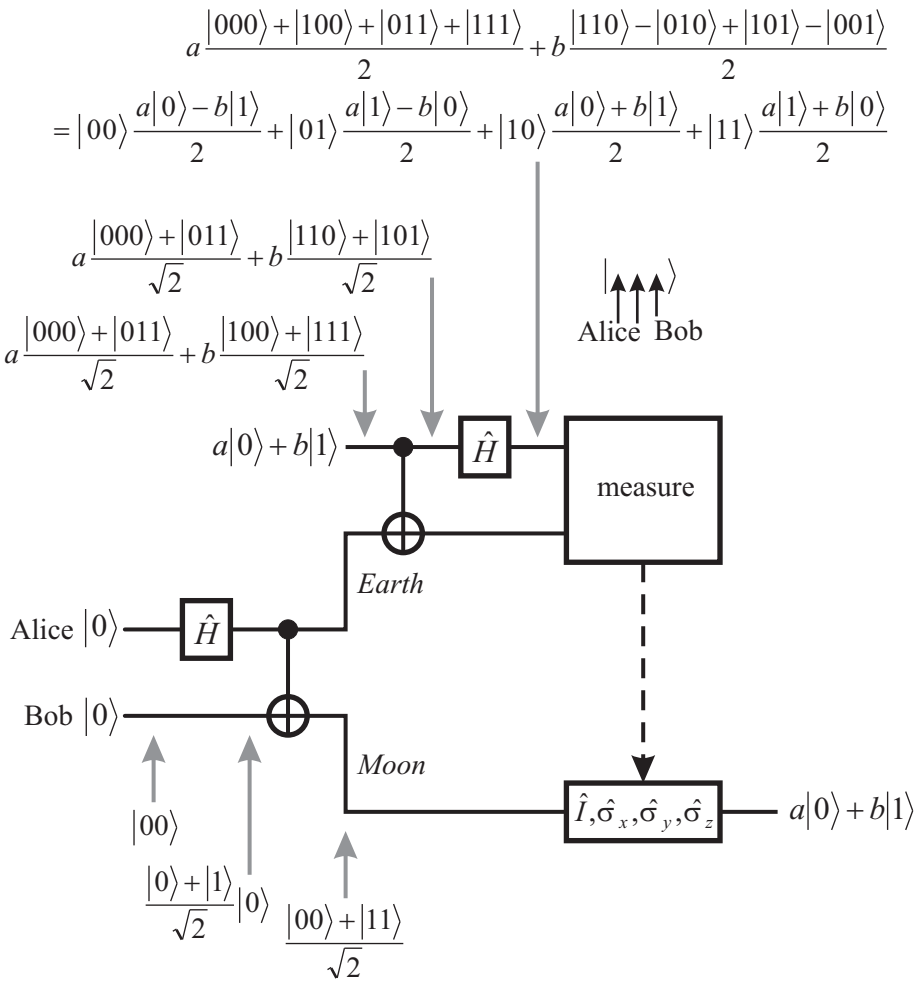

Figure 16.3. Quantum teleportation circuit.

The overall state after the second Hadamard transform can be written suggestively as shown. Through the entanglement, the coefficients of her qubit have moved over to Bob's qubit. When Alice measures her qubits she learns nothing about the value of a and b. But the outcome does force Bob's qubit into the state tagged by the result of her measurement. Alice can use a classical channel to communicate to Bob which of the four possible outcomes she obtained, and then Bob can use that information to change the

sign or swap the terms in his qubit as needed according to

$$
\begin{aligned}
|00\rangle &\rightarrow \hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\
|01\rangle &\rightarrow i\hat{\sigma}_y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\
|10\rangle &\rightarrow \hat{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
|11\rangle &\rightarrow \hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad .
\end{aligned} \tag{16.110}
$$

Once he's finished his qubit has the coefficients of the hers. Because quantum information can't be copied, this really *is* her qubit; the one she started out with was consumed by her measurement. At the instant it disappears its quantum information is transferred to Bob. For obvious reasons this has come to be called *teleportation* [Bennett *et al*., 1993]. While there's a long way to go before this is a practical way to reach your spaceship, it has been demonstrated experimentally [Furusawa *et al*., 1998].

16.3.4 Error Correction

Decoherence was once widely believed to be a terminal obstacle to practical applications of quantum information [Unruh, 1995]. After all, any attempt to measure the state of a quantum system in order to check for errors would necessarily disturb it. It was thus a great surprise to discover that not only could quantum errors be corrected, but it could be done using methods from Chapter 14.

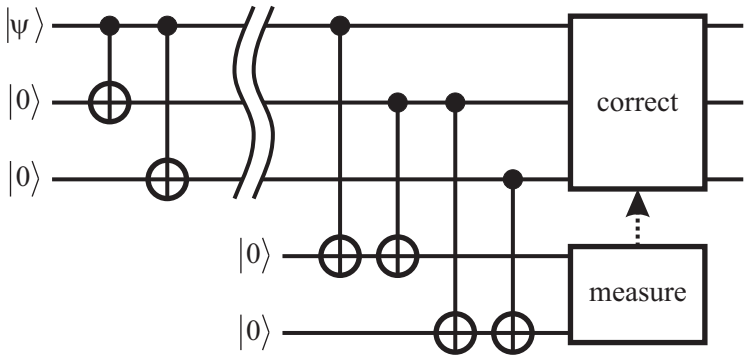

Figure 16.4. Quantum bit-flip error correction.

Figure 16.4 shows a simple scheme that protects a qubit $|\psi\rangle = a|0\rangle + b|1\rangle$ from bit flips. Like classical error correction, extra bits (here called *ancilla*) are introduced to add redundancy. Following the first two CNOTs, the qubit is encoded as

$$|\psi\rangle = a|000\rangle + b|111\rangle \quad . \tag{16.111}$$

After the possible introduction of errors, two more ancilla are used. The sequential CNOTs flip these if the first bit does not match the second, and if the second bit does not match

the third. Because of the entanglement, measurement of the state of these ancilla then forces the encoded qubits into one of these states.

Consider what happens if the first bit is flipped by an error

$$|\psi\rangle \to a|100\rangle + b|011\rangle \quad . \tag{16.112}$$

Following the ancilla measurement the first bit is found to disagree with the second and third. This can unambiguously be determined to be due to an error in the first bit, which can then be flipped back without learning anything about a or b. This is just a quantum version of a three-bit majority code.

This alone is not sufficient because qubits, unlike classical bits, have a phase as well as an amplitude. If there is a phase-flip error that changes the relative signs of the first qubit,

$$|\psi\rangle \to a|000\rangle - b|111\rangle \quad , \tag{16.113}$$

this cannot be recognized by the code and will erroneously appear as an error in the coefficient.

Phase-flip errors can be corrected by using Hadamards to encode the qubit in superpositions

$$\begin{aligned} |\psi\rangle = {} & a\left(|0\rangle + |1\rangle\right)\left(|0\rangle + |1\rangle\right)\left(|0\rangle + |1\rangle\right) \\ & + b\left(|0\rangle - |1\rangle\right)\left(|0\rangle - |1\rangle\right)\left(|0\rangle - |1\rangle\right) \quad . \end{aligned} \tag{16.114}$$

Now if there is a relative sign error in the first qubit

$$\begin{aligned} |\psi\rangle \to {} & a\left(|0\rangle - |1\rangle\right)\left(|0\rangle + |1\rangle\right)\left(|0\rangle + |1\rangle\right) \\ & + b\left(|0\rangle + |1\rangle\right)\left(|0\rangle - |1\rangle\right)\left(|0\rangle - |1\rangle\right) \end{aligned} \tag{16.115}$$

it can be recognized and corrected once again by checking to see if the pairs of encoded states match. But now a bit flip will cause a coefficient error

$$\begin{aligned} |\psi\rangle \to {} & a\left(|1\rangle + |0\rangle\right)\left(|0\rangle + |1\rangle\right)\left(|0\rangle + |1\rangle\right) \\ & + b\left(|1\rangle - |0\rangle\right)\left(|0\rangle - |1\rangle\right)\left(|0\rangle - |1\rangle\right) \\ = {} & a\left(|0\rangle + |1\rangle\right)\left(|0\rangle + |1\rangle\right)\left(|0\rangle + |1\rangle\right) \\ & - b\left(|0\rangle - |1\rangle\right)\left(|0\rangle - |1\rangle\right)\left(|0\rangle - |1\rangle\right) \quad . \end{aligned} \tag{16.116}$$

Both kinds of errors can be prevented by using both schemes, encoding the state in nine qubits as

$$\begin{aligned} |\psi\rangle = {} & a\left(|000\rangle + |111\rangle\right)\left(|000\rangle + |111\rangle\right)\left(|000\rangle + |111\rangle\right) \\ & + b\left(|000\rangle - |111\rangle\right)\left(|000\rangle - |111\rangle\right)\left(|000\rangle - |111\rangle\right) \quad . \end{aligned} \tag{16.117}$$

Now a bit flip

$$\begin{aligned} |\psi\rangle \to {} & a\left(|100\rangle + |011\rangle\right)\left(|000\rangle + |111\rangle\right)\left(|000\rangle + |111\rangle\right) \\ & + b\left(|100\rangle - |011\rangle\right)\left(|000\rangle - |111\rangle\right)\left(|000\rangle - |111\rangle\right) \end{aligned} \tag{16.118}$$

can be corrected within one of the encoded states, and a phase flip

$$\begin{aligned} |\psi\rangle \to {} & a\left(|000\rangle - |111\rangle\right)\left(|000\rangle + |111\rangle\right)\left(|000\rangle + |111\rangle\right) \\ & + b\left(|000\rangle + |111\rangle\right)\left(|000\rangle - |111\rangle\right)\left(|000\rangle - |111\rangle\right) \end{aligned} \tag{16.119}$$

can be corrected between them. An arbitrary quantum error can be written in terms of bit and phase flips, hence this scheme will repair any kind of damage to the qubit [Shor, 1995].

The possibility of quantum error correction was lurking in equation (14.33), which showed that in a classical block code the syndrome identifies an error without providing any information about the message. In quantum language this is called a *non-demolition measurement*. Through this insight much of the classical theory of error correction has been lifted into a quantum setting [Steane, 1996; Calderbank & Shor, 1996].

The importance of error correction extends far beyond communications. Because a computer or a storage device can be considered to be a channel, error correction is essential to their reliable operation. The existence of a channel capacity in these settings has the profound implication that as long as the physical error rate is below a threshold, arbitrarily long computations and storage times can be realized with imperfect components [von Neumann, 1956; Winograd & Cowan, 1963]. The same thing is true quantum mechanically in *fault-tolerant* quantum computing [Shor, 1996; Knill *et al*., 1998b].

Assume that a logical operation has a probability p of introducing an error. If instead the same operation is performed on qubits that have been encoded in an error-correction scheme that can fix n errors, the probability of failure after the subsequent decoding becomes Cp^{n+1}. The exponent is $n + 1$ because correction fails if there are that many errors, and C is a factor that accounts for the overhead of the extra steps needed for the encoding and decoding. If the error probability for a single step is below

$$Cp^{n+1} < p \;\Rightarrow\; p < C^{-1/n} \tag{16.120}$$

then the added work of the error correction is more than compensated for, and can be applied iteratively to improve the performance. A clever way to do this is with a *concatenated* code. This encodes qubits, which themselves are encoded qubits, which in turn are encoded qubits, and so forth. If one block encoding fails with a probability of Cp^{n+1}, the second level will fail if $n + 1$ blocks have an error, with a total probability of $C(Cp^{n+1}t)^{n+1}$, and a three-level encoding will fail with a probability of $C(C(Cp^{n+1})^{n+1})^{n+1}$. As long as the initial error is small enough, this provides a super-exponential improvement (with a corresponding increase in the number of extra qubits).

16.4 COMPUTATION

The quantum circuits we've considered so far have been used for coding, but because they can evaluate arbitrary functions it's natural to wonder whether they also have logical capabilities beyond their classical counterparts. The answer is a dramatic yes.

To understand what a computer can do, it's necessary to know something about the kinds of questions that can be asked of it. There are some problems that are known to be impossible to solve, like deciding if a computer program will halt without actually having to run it [Turing, 1936]. All of the remaining problems are theoretically soluble, but these solutions may not be practically usable given engineering constraints such as the length of a lifespan. While the execution time of a program will depend on the details of the machine used to run it, an essential distinction is between problems requiring a number of steps that is exponential versus polynomial in the problem size. For non-trivial questions the

size of an exponential will almost always exceed the amount of one's patience, hence these are usually considered to be intractable. Because the *Church–Turing thesis* assures us that reasonably-defined computers can execute each other's programs with a polynomial-time prefactor to run an emulator, the distinction between polynomial and exponential is invariant over a broad class of architectures.

Problems such as arithmetic or sorting that can be solved in a time that is polynomial in the problem size are said to belong to the class **P**. There is a larger class of problems that have solutions that can be checked in a polynomial time, but that might require longer than that to find an answer. This is the class **NP**. It seems apparent that $\mathbf{P} \neq \mathbf{NP}$ because the latter class could include many more problems, but proving this remains the greatest open problem in computer science [Garey & Johnson, 1979].

Within the class **NP** there is a subset of problems that are known to be as hard as any other one. If it's possible to efficiently solve any one of these **NP**-*complete* problems, then all of the other ones in **NP** can be solved efficiently. *Cook's Theorem* is the proof of this remarkable fact [Cook, 1971]. It applies to the *satisfiability* (SAT) problem of deciding whether there is an argument that satisfies a Boolean expression; other **NP**-complete problems include the traveling salesman problem of finding the shortest route connecting a group of cities, and coloring a graph with distinct colors on the vertices.

Efficient quantum algorithms have not been found for solving **NP**-complete problems, and there are some problems for which it's known that quantum mechanics provides no speed-up [Farhi *et al*., 1998]. While the possibility of finding an **NP**-complete algorithm remains an open question, there do exist good quantum algorithms for many important problems, including searching, factoring, and quantum simulation.

16.4.1 Searching

A very general way to formulate computational problems is in terms of an *oracle* $f(x)$ that equals 1 for arguments x that solve a problem of interest. The x might be entries in a database and $f(x)$ a test for answers to a query, or x could be a pair of integers and $f(x)$ a check for prime factors.

If there are N possible values of x, and they are unsorted and nothing is known about the inner workings of the oracle, then the only possible algorithm is to go through all arguments exhaustively to find an answer in $\mathcal{O}(N)$ steps. Think of finding a name in a phone book given a phone number, or randomly dialing a padlock to find the combination. Lov Grover astounded the world by showing that a quantum computer could do the problem in $\sqrt{N}$ steps [Grover, 1997; Grover, 1998].

Start by preparing a superposition of all possible logical states, which can be accomplished in a single step by applying a Hadamard transform to each of the qubits used to represent the states (Problem 16.6)

$$|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{x=1}^{N} |x\rangle \quad . \tag{16.121}$$

Now assume that there are M values of x for which $f(x) = 1$. Define $|1\rangle$ to be the

superposition of these states

$$|1\rangle = \frac{1}{\sqrt{M}} \sum_{f(x)=1} |x\rangle \tag{16.122}$$

and $|0\rangle$ to be what's left

$$|0\rangle = \frac{1}{\sqrt{N-M}} \sum_{f(x)=0} |x\rangle \quad , \tag{16.123}$$

so that

$$|\psi\rangle = \sqrt{\frac{N-M}{N}}\,|0\rangle + \sqrt{\frac{M}{N}}\,|1\rangle \quad . \tag{16.124}$$

If an angle θ is defined by

$$\sin\theta = \sqrt{\frac{M}{N}} \tag{16.125}$$

then the superposition state $|\psi\rangle$ can be rewritten in this basis as

$$|\psi\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} \quad . \tag{16.126}$$

Next define an operator

$$\begin{aligned} \hat{U}_1 &= \hat{I} - 2|1\rangle\langle 1| \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad . \end{aligned} \tag{16.127}$$

This flips the sign of the states that satisfy the oracle. While this might seem to require knowing the answer to the problem in advance, quantum mechanics allows the oracle to be applied to all arguments simultaneously. The problem is that flipping the signs alone is of no use because there are still N states to check to find the desired ones. To help uncover them, define a second operator

$$\begin{aligned} \hat{U}_\psi &= 2|\psi\rangle\langle\psi| - \hat{I} \\ &= 2\begin{bmatrix} \sin\theta\sin\theta & \sin\theta\cos\theta \\ \cos\theta\sin\theta & \cos\theta\cos\theta \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2\sin^2\theta - 1 & 2\sin\theta\cos\theta \\ 2\cos\theta\sin\theta & 2\cos^2\theta - 1 \end{bmatrix} \end{aligned} \tag{16.128}$$

that inverts the coefficients of an arbitrary state around their mean value. Now look what happens when these two opeartors are applied sequentially:

$$\begin{aligned} \hat{U}_\psi\hat{U}_1 &= \begin{bmatrix} 1 - 2\sin^2\theta & 2\sin\theta\cos\theta \\ -2\sin\theta\cos\theta & 2\cos^2\theta - 1 \end{bmatrix} \\ &= \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{bmatrix} \quad . \end{aligned} \tag{16.129}$$

This is just a rotation by $2\theta = 2\sqrt{M/N}$ in the $|0\rangle, |1\rangle$ space. $|\psi\rangle$ started out with an

angle of θ in that space due to the small fraction of states that satisfy the oracle. If $\hat{U}_\psi \hat{U}_1$ is applied iteratively I times, all of the probability will be in the $|1\rangle$ state after a rotation of $\pi/2$, requiring a number of iterations

$$\begin{aligned} \theta + I2\theta &= \frac{\pi}{2} \\ I &= \frac{\pi}{4}\frac{1}{\theta} - \frac{1}{2} \\ &= \frac{\pi}{4}\sqrt{\frac{N}{M}} - \frac{1}{2} \quad . \end{aligned} \tag{16.130}$$

After $\mathcal{O}(\sqrt{N})$ iterations the system will be in the $|x\rangle$ state with certainty, after which a single measurement can obtain an answer that satisfies the oracle!

If an answer could be found in $\log N$ time then this algorithm would efficiently solve **NP**-complete problems, but regrettably $\sqrt{N}$ is known to be a lower bound on the possible speed-up [Boyer *et al.*, 1998]. Nevertheless this can still be a very significant improvement, particularly given the broad applicability of oracle problems; searching a database of a million entries in a thousand steps is an enormous quantitative if not qualitative advantage.

16.4.2 Transforms and Factoring

The most significant result in quantum computing, and perhaps in all of computational complexity theory, was Peter Shor's proof that a quantum computer could find prime factors in polynomial time [Shor, 1997]. Factoring is one of the hardest problems that is not an **NP**-complete problem. Not only was it widely believed to require exponential time (Section 14.3.4), existing systems for secure electronic transactions rely on that belief. Unlike most results in the study of computational complexity that rest on problem classifications that remain open questions, here was a constructive demonstration of how to turn a problem from intractable to tractable.

Shor's result, and related algorithms, use a *Quantum Fourier Transform* (*QFT*). This is defined by applying a classical *Discrete Fourier Transform* (*DFT*) to the coefficients of a quantum state

$$\sum_{x=0}^{N-1} a_n |x\rangle \;\mapsto\; \sum_{k=0}^{N-1}\sum_{n=0}^{N-1} a_n e^{2\pi i k x/N} |k\rangle \quad . \tag{16.131}$$

A classical DFT requires $\mathcal{O}(N^2)$ steps because of the double sum over N indices. A *Fast Fourier Transform* (*FFT*) reduces this to $\mathcal{O}(N \log N)$ by recognizing that $e^{2\pi i k x/N}$ does not need to be evaluated for values of kx/N that are a power of 2. To turn this into an algorithm (quantum or classical), note that because x and k are integer indices they can be written in a binary expansion with coefficients that can be either 0 or 1

$$\begin{aligned} x &= x_0 + 2^1 x_1 + \cdots + 2^{n-1} x_{n-1} \\ k &= k_0 + 2^1 k_1 + \cdots + 2^{n-1} k_{n-1} \end{aligned} \tag{16.132}$$

where $N = 2^n$. Then discarding all of the terms in xk with an exponent of 2^n or higher

leaves the non-trivial terms

$$\frac{kx}{2^n} = k_{n-1}\left(2^{-1}x_0\right) + k_{n-2}\left(2^{-1}x_1 + 2^{-2}x_0\right) + \cdots \quad . \tag{16.133}$$

This expansion can be plugged back into the definition of the QFT to find its action on a basis vector

$$\begin{aligned} |x\rangle &= |x_0 x_1 \cdots x_{n-1}\rangle \\ &\mapsto \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i k x/N} |k\rangle \\ &= \frac{1}{2^{n/2}} \sum_{k_{n-1}=0}^{1} \sum_{k_{n-2}=0}^{1} e^{2\pi i\left[k_{n-1}\left(2^{-1}x_0\right) + k_{n-2}\left(2^{-1}x_1 + 2^{-2}x_0\right) + \cdots\right]} |k_{n-1}\cdots k_1 k_0\rangle \\ &= \frac{1}{2^{n/2}} \left(|0\rangle + e^{2\pi i\left(2^{-1}x_0\right)}|1\rangle\right)\left(|0\rangle + e^{2\pi i\left(2^{-1}x_1 + 2^{-2}x_0\right)}|1\rangle\right) + \cdots \quad . \end{aligned} \tag{16.134}$$

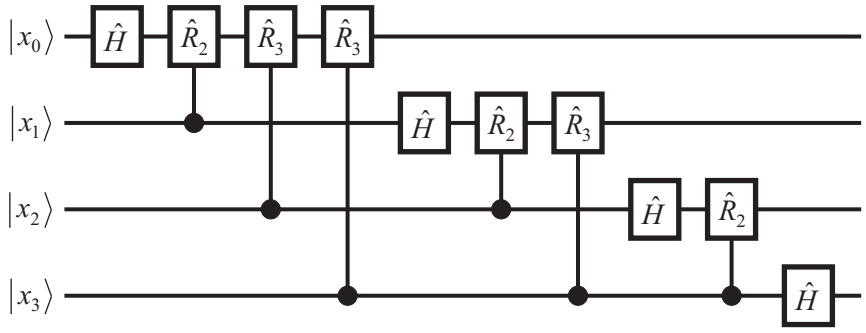

Figure 16.5. Circuit for a quantum Fourier transform.

The product representation in equation (16.134) has a natural implementation in a quantum circuit, shown in Figure 16.5. Input is the state $|x_0 x_1 \cdots x_{n-1}\rangle$. After the first Hadamard this becomes

$$\left(|0\rangle + e^{2\pi i\left(2^{-1}x_0\right)}|1\rangle\right)|x_1 \cdots x_{n-1}\rangle \tag{16.135}$$

because the sign in front of the $|1\rangle$ is a + if $x_0 = 0$ and a $-$ if $x_0 = 1$ (equation 16.108). Then comes a conditional phase shift defined by

$$\hat{R}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{2\pi i/2^k} \end{bmatrix} \quad . \tag{16.136}$$

This does nothing to the $|0\rangle$ state, but adds an extra phase shift to the $|1\rangle$ state based on the value of x_1, leaving

$$\left(|0\rangle + e^{2\pi i\left(2^{-1}x_0 + 2^{-2}x_1\right)}|1\rangle\right)|x_1 \cdots x_{n-1}\rangle \quad . \tag{16.137}$$

After applying all of the conditional phase shifts, and then the Hadamard to the next bit,

the state is

$$\left(|0\rangle + e^{2\pi i(2^{-1}x_0+2^{-2}x_1+\cdots)}|1\rangle\right)\left(|0\rangle + e^{2\pi i(2^{-1}x_1)}|1\rangle\right)|x_2\cdots x_{n-1}\rangle \quad . \tag{16.138}$$

Continuing in this manner reproduces equation (16.134).

One Hadamard and $n-1$ phase gates are applied to $|x_0\rangle$ for a total of n operations, then there are $n-1$ gates for $|x_1\rangle$, and so forth, adding up to a total of $n+(n-1)+\cdots+1 = n(n+1)/2$ steps. This is for a single basis vector, but because of the linearity of quantum mechanics it will also work on an arbitrary superposition, therefore the QFT reduces the total number of steps to $\mathcal{O}\left(n^2\right) = \mathcal{O}\left((\log N)^2\right)$, quite a savings over the FFT. While this is not directly applicable to classical signal-processing problems because of the difficulty of loading in and reading out the data, it can be used as a primitive in a number of quantum algorithms, most notably factoring.

The QFT appears in Shor's algorithm as a way to use quantum interference for period-finding:

- Given an integer N to factor, pick a power of two $K = 2^L$ such that $N^2 < K < 2N^2$.
- Choose random integer $x < N$.
- Initialize a pair of K-qubit registers

$$\frac{1}{\sqrt{K}}\sum_{k=0}^{K-1}|0\rangle|0\rangle \quad . \tag{16.139}$$

- Create the uniform superposition in the first register

$$\frac{1}{\sqrt{K}}\sum_{k=0}^{K-1}|k\rangle|0\rangle \quad . \tag{16.140}$$

- Perform a *modular exponentiation* of the first register on the second

$$\frac{1}{\sqrt{K}}\sum_{k=0}^{K-1}|k\rangle|x^k \bmod N\rangle \quad . \tag{16.141}$$

- Make a measurement on the second register. If the result is

$$x^k \bmod N = y \tag{16.142}$$

  then the first register will collapse into those values of x compatible with y. Because the modular exponentiation is periodic in k with an *order* r [Koblitz, 1994], this leaves the first register in the state

$$\frac{1}{\sqrt{M}}\sum_{m=0}^{M-1}|l+mr\rangle \quad , \tag{16.143}$$

  where l is an offest that depends on y, and M is the number of values of k satisfying equation (16.142).
- To find r, perform a QFT

$$\frac{1}{\sqrt{M}}\sum_{m=0}^{M-1}\frac{1}{\sqrt{K}}\sum_{n=0}^{K-1}e^{2\pi i(l+mr)n/K}|n\rangle \quad . \tag{16.144}$$

- Measure the first register. The probability of obtaining a value n is proportional to its coefficient

$$\left|\sum_{m=0}^{M-1} e^{2\pi i(l+mr)n/K}\right|^2 = \left|\sum_{m=0}^{M-1} e^{2\pi imrn/K}\right|^2 \quad . \tag{16.145}$$

  Since the terms in the sum will add incoherently unless rn/K is an integer, the result is $n = KC/r$, where C is a random integer.
- From KC/r find the order r, from which the factors of N can be found [Ekert & Jozsa, 1996].

A complete implementation of this algorithm requires $\mathcal{O}\left((\log N)^3\right)$ gates, dominated by the calculation of the modular exponentiation [Beckman *et al*., 1996].

16.4.3 Simulation

One of the original questions that inspired the study of quantum information was whether a quantum computer could efficiently simulate the evolution of another quantum system [Feynman, 1982]. If, for example, an N-spin system is being studied, modeling its state requires a state vector with 2^N elements, and operators on it are $2^N \times 2^N$ matrices. Since $2^{40} = 10^{12}$, simulating just 40 spins would exceed the capacity of the largest digital computers.

Not all quantum systems can be modeled efficiently on a quantum computer, but a tremendous simplification is possible if the Hamiltonian can be written as a sum [Lloyd, 1996]

$$e^{i\hat{\mathcal{H}}t/\hbar} = e^{i\sum_n \hat{\mathcal{H}}_n t/\hbar} \quad . \tag{16.146}$$

This will be the usual case if there are short-range interactions that extend near neighbors. We then want to find a numerical method to appoximate equation (16.146) on a quantum computer, but unlike classical numerical methods [Gershenfeld, 1999a] this one must recognize that the Hamiltonian terms in general won't commute.

For two terms, the evolution over a time Δt is to second order

$$\begin{aligned} & e^{i(\hat{A}+\hat{B})\Delta t} \\ &= \hat{1} + i(\hat{A}+\hat{B})\Delta t - \frac{1}{2}(\hat{A}+\hat{B})(\hat{A}+\hat{B})\Delta t^2 + \mathcal{O}(\Delta t^3) \\ &= \hat{1} + i\hat{A}\Delta t - \frac{1}{2}\hat{A}^2\Delta t^2 + i\hat{B}\Delta t - \frac{1}{2}\hat{B}^2\Delta t^2 + \frac{1}{2}\hat{A}\hat{B}\Delta t^2 + \frac{1}{2}\hat{B}\hat{A}\Delta t^2 + \mathcal{O}(\Delta t^3) \quad . \end{aligned} \tag{16.147}$$

An obvious guess would be to just apply each evolution separately, but this is incorrect at second order

$$\begin{aligned} & e^{i\hat{A}\Delta t}e^{i\hat{B}\Delta t} \\ &= \left(\hat{1} + i\hat{A}\Delta t - \frac{1}{2}\hat{A}^2\Delta t^2 + \mathcal{O}(\Delta t^3)\right)\left(\hat{1} + i\hat{B}\Delta t - \frac{1}{2}\hat{B}^2\Delta t^2 + \mathcal{O}(\Delta t^3)\right) \\ &= \hat{1} + i\hat{A}\Delta t - \frac{1}{2}\hat{A}^2\Delta t^2 + i\hat{B}\Delta t - \frac{1}{2}\hat{B}^2\Delta t^2 - \hat{A}\hat{B}\Delta t^2 + \mathcal{O}(\Delta t^3) \quad . \end{aligned} \tag{16.148}$$

That can be fixed by splitting the first operator as

$$\begin{aligned} & e^{i\hat{A}\Delta t/2}e^{i\hat{B}\Delta t}e^{i\hat{A}\Delta t/2} \\ &= \left(\hat{1} + \frac{i}{2}\hat{A}\Delta t - \frac{1}{8}\hat{A}^2\Delta t^2 + \mathcal{O}(\Delta t^3)\right)\left(\hat{1} + i\hat{B}\Delta t - \frac{1}{2}\hat{B}^2\Delta t^2 + \mathcal{O}(\Delta t^3)\right) \\ &\quad \left(\hat{1} + \frac{i}{2}\hat{A}\Delta t - \frac{1}{8}\hat{A}^2\Delta t^2 + \mathcal{O}(\Delta t^3)\right) \qquad (16.149) \\ &= \hat{1} + i\hat{A}\Delta t - \frac{1}{2}\hat{A}^2\Delta t^2 + i\hat{B}\Delta t - \frac{1}{2}\hat{B}^2\Delta t^2 - \frac{1}{2}\hat{A}\hat{B}\Delta t^2 - \frac{1}{2}\hat{B}\hat{A}\Delta t^2 + \mathcal{O}(\Delta t^3) \quad . \end{aligned}$$

Asymptotically, this leads to the *Trotter formula*

$$\lim_{N\to\infty}\left(e^{i\hat{A}t/N}e^{i\hat{B}t/N}\right)^N = e^{i(\hat{A}+\hat{B})t} \quad . \qquad (16.150)$$

The second-order cancellation can also be obtained by grouping the terms in the *Campbell–Baker–Hausdorf formula*

$$e^{i(\hat{A}+\hat{B})\Delta t} = e^{i\hat{A}\Delta t}e^{i\hat{B}\Delta t}e^{-[\hat{A},\hat{B}]\Delta t^2/2} + \mathcal{O}(\Delta t^3) \quad . \qquad (16.151)$$

These relationships are the basis of quantum simulation. Quantum circuits are devised that emulate the local Hamiltonian, and then these are applied iteratively to model the global evolution. A linear number of applications corresponds to an exponential amount of work in the Hilbert space of the quantum computer, which can be a physical system unrelated to the one being modeled [Somaroo *et al*., 1999].

Although quantum simulation has received less attention than other quantum algorithms because prospective users are people who already study quantum mechanics, it offers an exponential improvement with broad applicability to many areas of technological interest such as molecular structure and quantum effects in VLSI scaling.

16.4.4 Experimental Implementation

This is a fitting final section. Quantum computing is one of the most exciting opportunities in all of information processing, both for its fundamental implications and practical applications. The explanation of how these devices function crosses the traditional division between the theory of computation and physical theory, and experimental implementations draw on ideas from almost every page in this book.

Of all the candidate technologies to continue scaling beyond the VLSI era, quantum logic has one unique feature. Moore's Law is exponential; any classical approach demands an exponential increases in space or time. Even the Avogadro's number of elements in a DNA computer is quickly limited by the size of an exponential problem [Adleman, 1994]. Quantum computing accesses Hilbert space, the one exponential resource that has been untapped for computation. What's so remarkable about algorithms such as Shor's and Grover's is that they provide answers with certainty, rather than a probability of success that decreases inversely with the problem size.

Early theoretical studies of quantum computing used "designer" Hamiltonians with desired properties. The development of experimental approaches began with the recognition that physically-plausible local interactions were adequate for realizing arbitrary quantum algorithms [Lloyd, 1993]. Initial attempts to apply this insight ran afoul of the

essential paradox in quantum logic: the system must be protected from the external environment to remain quantum, but must remain accessible to the external environment for I/O and programming. Most physical systems fit into one of two categories, those that are externally isolated and internally uncoupled and have quantum coherence times that can be on the order of the age of the universe, and strongly-coupled interacting ones with subnanosecond coherence times.

Heroic experimental efforts are beginning to develop controllable quantum nonlinearities. Among the candidate systems are ions in laser-addressed optical traps [Cirac & Zoller, 1995; Wieman *et al*., 1999], atoms in optical cavities small enough to feel quantum electrodynamic effects [Ye *et al*., 1999], dilute spins in semiconductors with exchange interactions controlled by carriers induced by a gate voltage [Kane, 1998], and superconducting current loops [Mooij *et al*., 1999].

Prospects for their practical implementation improved significantly with the development of quantum error correction and fault-tolerant circuit design that can use imperfect components. While quantitative estimates depend on device and algorithm details, the required relative loss of coherence per gate for steady-state operation is in the experimentally accessible range of 10^{-3}–10^{-6} [Knill *et al*., 1998b].

A second insight that significantly simplifies the device design is to recognize that a macroscopic ensemble of microscopic quantum computers can be used. By representing each logical qubit in many physical qubits, intentional and unintentional external interactions become *weak measurements* that can paradoxically obtain continuous information about the state of the logical qubits without projecting them. Post-processing can be added to quantum algorithms so that a classical channel can be used to obtain quantum results from an ensemble [Gershenfeld & Chuang, 1997].

Nuclear spins in molecular liquids are screened and protected by rapid tumbling averaging out inter-molecular forces, giving coherence times on the order of seconds. Their exchange interactions through bonds provide a nonlinearity that can be controlled with RF fields. And mature chemical synthesis techniques can be used to produce macroscopic samples of molecules of interest. These desirable attributes led to the development of nuclear magnetic resonance pulse sequences for computation [Cory *et al*., 1997; Chuang *et al*., 1998a].

Two spins in a strong magnetic field will have a Hamiltonian [Ernst *et al*., 1994]

$$\hat{\mathcal{H}} = \hbar\omega_A \frac{\hat{\sigma}_A}{2} + \hbar\omega_B \frac{\hat{\sigma}_B}{2} + \hbar\omega_{AB} \frac{\hat{\sigma}_A}{2} \frac{\hat{\sigma}_B}{2} \quad . \tag{16.152}$$

ω_A and ω_B are the precession frequencies of the spins in the magnetic field, possibly differing because of a *chemical shift* in the local field strength. The ω_{AB} term represents nonlinear coupled evolution due to the exchange interaction. Free precession under this Hamiltonian corresponds to linear and nonlinear rotations, which along with rotation by transverse RF pulses provides the operators needed for univeral quantum circuits. These were used for the first demonstrations of non-trivial quantum algorithms [Chuang *et al*., 1998b; Jones & Mosca, 1998] and error correction [Cory *et al*., 1998].

Because the *Zeeman splitting* between spin eigenstates due to dipole coupling to a magnetic field is 10^{-5} smaller than kT at room temperature in fields of a few tesla, thermal spins are in a weakly-polarized high-temperature limit. Even though their evolution is

quantum, the pure states needed for algorithm initial conditions cannot be obtained from an equilibrium distribution (equation 16.98) by using unitary RF pulses. NMR techniques simulate non-unitary operators by adding extra degrees of freedom (spins, space, or time [Knill *et al*., 1998a]), but these do not scale beyond small demonstrations because of an exponential loss in signal strength as qubits are added due to the partition function denominator in equation (16.98) [Schack & Caves, 1999]. Overcoming this will require order unity spin polarization, which has been accomplished in simpler atomic systems by *optical pumping* [Song *et al*., 1999].

Whether or not any one of these schemes can eventually be scaled to defeat number-theoretic cryptosystems, their development is already having a significant impact on the study of both computation and physical systems. For many years it's been clear that computers can and will be constructed on atomic scales, the domain of *nanotechnology* [Feynman, 1992; Drexler, 1992; Merkle, 1998]. The demonstration by NMR of quantum algorithms in nuclear spin evolution shows that not only is this possible, but that nature is already a very powerful computer if it is interrogated in the right way, and that the kinds of computations it can perform go far beyond what can be conceived of by our classical intuition alone.

The interchange between the study of nature and computation increasingly operates in both directions. Physical theory has provided improved devices to compute with, but computation is also providing an improved language to described devices with. By viewing natural systems in a computational framework it can be possible to understand how to "program" them to obtain desired behavior. An example is the traditional difficulty in NMR of exchanging the quantum coefficients between a spin of interest and a more sensitive accessible one; it's been shown that this can be accomplished by a pulse sequence that implements the circuit for the `SWAP` operation (Problem 16.7) [Linden *et al*., 1999]. Another example is writing an arbitrary wave function into an atom by putting a pump laser in a feedback loop with a computer running a machine learning algorithm [Weinacht *et al*., 1999]. Such cross-fertilization could equally well be called *The Information of Physics Technology*, enhancing our understanding of, and ability to shape, the world around us by merging the descriptions of the information in a system with that of its physical properties.

16.5 SELECTED REFERENCES

[Baym, 1973] Baym, Gordon. (1973). *Lectures on Quantum Mechanics*. Reading: W.A. Benjamin.

A good intuitive introduction to quantum mechanics.

[Peres, 1993] Peres, Asher. (1993). *Quantum Theory: Concepts and Methods*. Boston: Kluwer Academic.

Modern quantum theory.

[Balian, 1991] Balian, Roger. (1991). *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics*. New York: Springer–Verlag. Translated by D. ter Haar and J.F. Gregg, 2 volumes.

Statistical mechanics with a strong quantum flavor.

[quant-ph] `http://xxx.lanl.gov`

The quantum physics e-print archive, where most results first appear.

[Nielsen & Chuang, 2000] Nielsen, M.A., & Chuang, I.L. (2000). *Quantum Computation and Quantum Information.* New York: Cambridge University Press.

A monumental compendium of most everything there is to know about quantum mechanics, and computation.

16.6 PROBLEMS

(16.1) How is Ehrenfest's theorem changed if the observable has an explicit time dependence?

(16.2) Show that that $\mathrm{Tr}(\hat{\rho}^2) \leq 1$.

(16.3) (*a*) Using the Pauli matrices in the z eigenbasis of equations (16.71), find the eigenvectors for $\hat{\sigma}_x$ and $\hat{\sigma}_y$.

(*b*) Using the result of the previous problem, apply to the singlet state $|\psi\rangle = \left(|\uparrow\rangle_1|\downarrow\rangle_2 - |\downarrow\rangle_1|\uparrow\rangle_2\right)/\sqrt{2}$ a projector $|m_x = 1/2\rangle_1\langle m_x = 1/2|_1$ onto the $m_{1x} = 1/2$ state of the first spin. What state is the second spin left in? Repeat this in the y direction.

(*c*) How are the products of eigenvalues $m_{1x}m_{2y}$ and $m_{1y}m_{2x}$ related?

(*d*) Recognizing that $\hat{\vec{\sigma}}_1$ commutes with $\hat{\vec{\sigma}}_2$, what does the last result imply for the relationship between $\langle\psi|\hat{\sigma}_{1x}\hat{\sigma}_{2y}|\psi\rangle$ and $\langle\psi|\hat{\sigma}_{1y}\hat{\sigma}_{2x}|\psi\rangle$?

(*e*) In the z eigenbasis, work out the tensor products to evaluate $\langle\psi|\hat{\sigma}_{1x}\hat{\sigma}_{2y} + \hat{\sigma}_{1y}\hat{\sigma}_{2x}|\psi\rangle$.

(*f*) Compare the results of the last two parts. What happened?

(16.4) Using the spectral representation, work out the rotation operators associated with the Pauli spin matrices.

(16.5) Find the matrix representation of the $\sqrt{\texttt{NOT}}$ gate that when applied twice gives a `NOT` gate (don't worry about getting the signs of the final state right).

(16.6) Show that applying Hadamard transformations individually to N qubits each in the $|0\rangle$ state puts them into an equal superposition of the 2^N possible logical states.

(16.7) What does this circuit do?