Phase Estimation

In this lecture we will describe Kitaev's phase estimation algorithm, and use it to obtain an alternate derivation of a quantum factoring algorithm. We will also use this technique to design quantum circuits for computing the Quantum Fourier Transform modulo an arbitrary positive integer.

0.1 Phase Estimation Technique

In this section, we define the phase estimation problem and describe an efficient quantum circuit for it.

Property 0.1 *Let U be a $N \times N$ unitary transformation. U has an orthonormal basis of eigenvectors $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_N\rangle$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$, where $\lambda_j = e^{2\pi i \theta_j}$ for some θ_j.*

Proof: U, being unitary, maps unit vectors to unit vectors and hence all the eigenvalues have unit magnitude, i.e. they are of the form $e^{2\pi i\theta}$ for some θ. Let $|\psi_j\rangle$ and $|\psi_k\rangle$ be two distinct eigenvectors with distinct eigenvalues λ_j and λ_k. We have that $\lambda_j \langle \psi_j, \psi_k\rangle = \langle \lambda_j \psi_j, \psi_k\rangle = \langle U\psi_j, \psi_k\rangle = \langle \psi_j, U\psi_k\rangle = \langle \psi_j, \lambda \psi_k\rangle = \lambda_k \langle \psi_j, \psi_k\rangle$. Since $\lambda_j \neq \lambda_k$, the inner product $\langle \psi_j, \psi_k\rangle$ is 0, i.e. the eigenvectors $|\psi_j\rangle$ and $|\psi_k\rangle$ are orthonormal. □

Given a unitary transformation U, and one of its eigenvector $|\psi_j\rangle$, we want to figure out the corresponding eigenvalue λ_j (or, equivalently, θ_j). This is the phase estimation problem.

Definition 0.2 *For any unitary transformation U, let C-U stand for a "controlled U" circuit which conditionally transforms $|\psi\rangle$ to $U|\psi\rangle$ as shown in Figure* **??**.

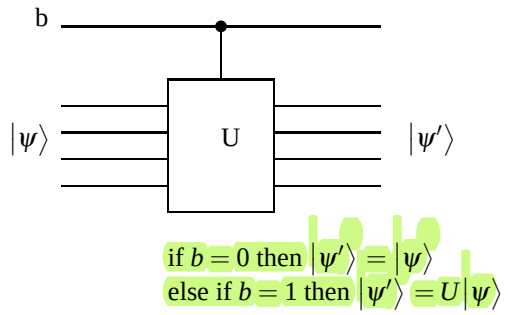

Figure 0.1: Controlled U Circuit

Assume that we have a circuit which implements the controlled U transformation (We will see later in the course how to construct a circuit that implements a controlled U transformation given a circuit that implements U). The phase estimation circuit in Figure **??** can be used to estimate the value of θ.

The phase estimation circuit performs the following sequence of transformations:

$$\begin{aligned} |0\rangle|\psi\rangle &\xrightarrow{H} \text{s.t. } (|0\rangle + |1\rangle)|\psi\rangle \\ &\xrightarrow{C\text{-}U} \text{s.t. } |0\rangle|\psi\rangle + \text{s.t. } |1\rangle\lambda|\psi\rangle \\ &= (\text{ s.t. } |0\rangle + \frac{\lambda}{\sqrt{2}}|1\rangle) \otimes |\psi\rangle \end{aligned}$$

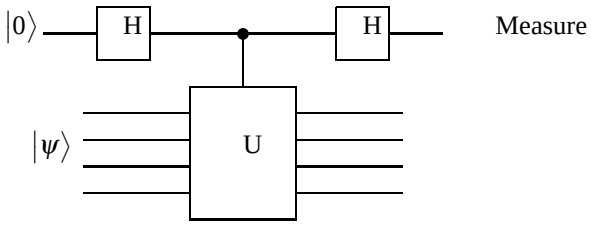

Figure 0.2: Phase Estimation Circuit

Note that after the *C-U* transformation, the eigenvector remains unchanged while we have been able to put λ into the phase of the first qubit. A Hadamard transform on the first qubit will transform this information into the amplitude which we will be able to measure.

$$\xrightarrow{H} \quad \frac{1+\lambda}{\sqrt{2}}|0\rangle + \frac{1-\lambda}{\sqrt{2}}|1\rangle$$

Let $P(0)$ and $P(1)$ be the probability of seeing a zero and one respectively on measuring the first qubit. If we write $\lambda = e^{2\pi i\theta}$, we have:

$$P(0) = \left|\frac{1+\cos 2\pi\theta + i\sin 2\pi\theta}{\sqrt{2}}\right|^2 = \frac{1+\cos 2\pi\theta}{2}$$

$$P(1) = \left|\frac{1-\cos 2\pi\theta - i\sin 2\pi\theta}{\sqrt{2}}\right|^2 = \frac{1-\cos 2\pi\theta}{2}$$

There is a bias of $\frac{1}{2}\cos 2\pi\theta$ in the probability of seeing a 0 or 1 upon measurement. Hence, we can hope to estimate θ by performing the measurement several times. However, to estimate $\cos 2\pi\theta$ within m bits of accuracy, we need to perform $\Omega(2^m)$ measurements. This follows from the fact that estimating the bias of a coin to within ε with probability at least $1-\delta$ requires $\Theta(\frac{\log(1/\delta)}{\varepsilon^2})$ samples.

We will now see how to estimate θ efficiently. Suppose we can implement the C_m-U transformation as defined below.

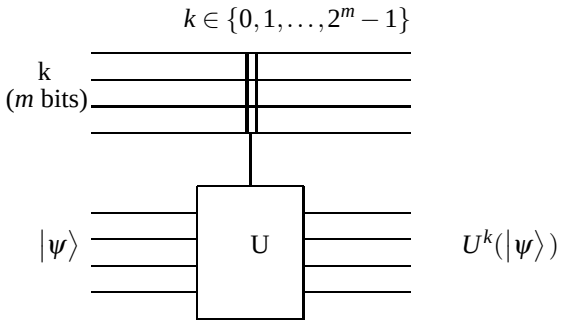

Figure 0.3: m-Controlled U Circuit

Definition 0.3 *For any unitary transformation U, let C_k-U stand for a "k-controlled U" circuit which implements the transformation $|k\rangle \otimes |\psi\rangle \longrightarrow |k\rangle \otimes U^k|\psi\rangle$ as shown in Figure* **??**.

Estimating θ within m bits of accuracy is equivalent to estimating integer j, where $\frac{j}{2^m}$ is the closest approximation to θ. Let $M = 2^m$ and $w_M = e^{\frac{2\pi i}{M}}$.

The circuit in Figure **??** performs the following sequence of transformations:

$$\begin{aligned} |0^m\rangle|\psi\rangle &\xrightarrow{H^{\otimes m}} \left(\frac{1}{\sqrt{M}}\sum_{k=0}^{M-1}|k\rangle\right)\otimes|\psi\rangle \\ &\xrightarrow{C_m\text{-}U} \left(\frac{1}{\sqrt{M}}\sum_{k=0}^{M-1}\lambda^k|k\rangle\right)\otimes|\psi\rangle \\ &= \left(\frac{1}{\sqrt{M}}\sum_{k=0}^{M-1}w_M^{jk}|k\rangle\right)\otimes|\psi\rangle \end{aligned}$$

Note that the first register now contains the Fourier Transform mod M of j and if we apply the reverse of the Fourier Transform mod M (note that quantum circuits are reversible), we will get back j.

$$\xrightarrow{QFT_M^{-1}} |j\rangle\otimes|\psi\rangle$$

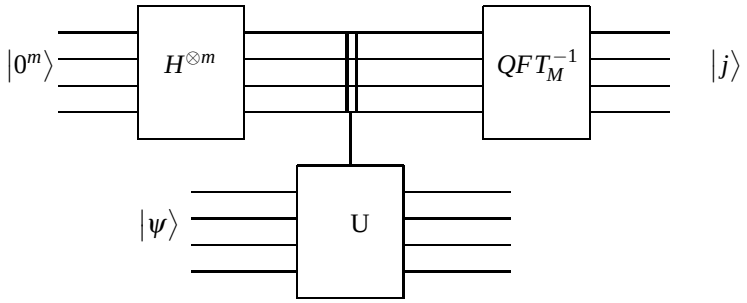

Figure 0.4: Efficient Phase Estimation Circuit

If $\theta = \frac{j}{2^m}$, then clearly the circuit outputs j. If $\theta \approx \frac{j}{2^m}$, then the circuit outputs j with high probability (Exercise!).

0.2 Kitaev's Factoring Algorithm

In this section, we will see how to use the phase estimation circuit to factor a number.

Recall that the problem of factoring reduces to the problem of order finding. To factor N, it is sufficient to pick a random number a and compute the minimum positive r such that $a^r \equiv 1 \bmod N$. With reasonable probability, r is even and $a^{r/2} \not\equiv \pm \bmod N$ and hence $N \mid a^r - 1$, i.e. $N \mid (a^{r/2}+1)(a^{r/2}-1)$. Since N does not divide $a^{r/2} \pm 1$, it must be the case that a part of it divides $a^{r/2}+1$ and hence $gcd(N, a^{r/2}+1)$ is a non-trivial factor of N.

We now reduce the problem of order finding to the phase estimation problem. Consider the unitary transformation

$M_a : |x\rangle \rightarrow |xa \bmod N\rangle$. Its eigenvectors are $|\psi_k\rangle = \frac{1}{\sqrt{r}}\left(|1\rangle + w^{-k}|a\rangle + \ldots + w^{-k(r-1)}|a^{r-1}\rangle\right)$, where $w = e^{2\pi i/r}$:

$$\begin{aligned} M_a|\psi_k\rangle &= \frac{1}{\sqrt{r}}\left(|a\rangle + w^{-k}|a^2\rangle + \ldots + w^{-k(r-1)}|a^r\rangle\right) \\ &= w^k \frac{1}{\sqrt{r}}\left(|1\rangle + w^{-k}|a\rangle + \ldots + w^{-k(r-1)}|a^{r-1}\rangle\right) \\ &= w^k|\psi_k\rangle \end{aligned}$$

It follows that $|\psi_k\rangle$ is an eigenvector of M_a with eigenvalue w^k. Hence, if we can implement the C_m-M_a transformation and construct the eigenvector ψ_k for some suitable k, we can use the phase estimation circuit to obtain an approximation to the eigenvalue w^k and therefore reconstruct r as follows: $w^k = e^{2\pi i\theta}$ for $\theta = k/r$. Recall that phase estimation reconstructs $\theta \approx \frac{j}{2^m}$ where j is the output of the phase estimation procedure carried out to m bits of precision. Thus with high probability $\frac{j}{2^m}$ is a very close approximation to $\frac{k}{r}$. Assuming that k is relatively prime to r (which we will ensure with high probability) we can estimate r using the method of continued fractions if we choose $M \approx N^2$.

Lets look carefully at the C_m-M_a transformation. It transforms $|k\rangle|x\rangle \rightarrow |xa^k \bmod N\rangle$. But this is precisely the transformation that does modular exponentiation. There exists a classical circuit that performs this transformation in $O(|x|^2|k|)$ time, and thus we can construct a quantum circuit that implements the C_m-M_a transformation.

It is not obvious how to obtain an eigenvector $|\psi_k\rangle$ for some k, but it is easy to obtain the uniform superposition of the eigenvectors $|\psi_0\rangle, |\psi_1\rangle \ldots |\psi_{r-1}\rangle$. Note that $\frac{1}{\sqrt{r}}\sum_{k=0}^{r-1}|\psi_k\rangle = |1\rangle$. Hence, if we use $|1\rangle$ as the second input to the phase estimation circuit, then we will be able to measure a random eigenvalue w^k, where k is chosen u.a.r. from the set $\{0,\ldots,r-1\}$. Note that $k = 0$ is completely useless for our purposes. But k will be relatively prime to r with reasonable probability.

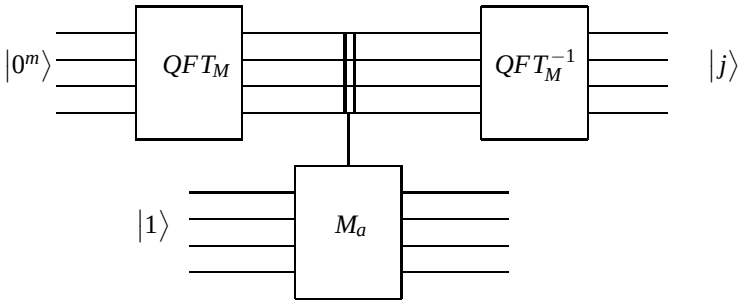

Figure 0.5: Order Finding Circuit (Kitaev's)

With these observations, it is easy to see that the circuit in Figure **??** outputs $|j\rangle$ with high probability, where $\frac{j}{2^m}$ is the closest approximation to $\frac{k}{r}$ for some random k. Note that with reasonable probability, k is relatively prime to r and if that is the case, then we can estimate r using the method of continued fractions if we choose $M \approx N^2$. Note also that QFT_M and $H^{\otimes q}$ act in an identical manner on $|0^q\rangle$, so $H^{\otimes q}$ could be used in the above circuit in place of the QFT_M transformation.

Though the thinking and the analysis behind the Kitaev's and Shor's order-finding algorithm are different, it is interesting to note that the two circuits are almost identical. Figure **??** describes the Shor's circuit. The quantities q, Q and x in the Shor's algorithm correspond to m, M and a in the Kitaev's algorithm. Also, note that raising a to some power is same as performing controlled multiplication.

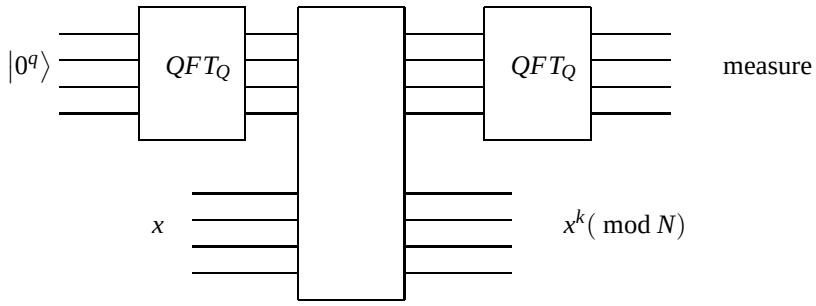

Figure 0.6: Order Finding Circuit (Shor's)

0.3 QFT mod Q

In this section, we will present Kitaev's quantum circuit for computing Fourier Transform over an arbitrary positive integer Q, not necessarily a power of 2. Let m be such that $2^{m-1} < Q \leq 2^m$ and let $M = 2^m$.

Recall that the Fourier Transform mod Q sends

$$|a \bmod Q\rangle \longrightarrow \frac{1}{\sqrt{Q}} \sum_{b=0}^{Q-1} w^{ab} |b\rangle \stackrel{\text{let}}{=} |\chi_a\rangle$$

where $w = e^{2\pi i/Q}$. Note that $\{|\chi_a\rangle \mid a = 0,1,\ldots,Q-1\}$ forms an orthonormal basis, so we may regard the Fourier Transform as a change of basis.

Consider the following sequence of transformations, which computes something close to the Fourier Transform mod Q:

$$|a\rangle|0\rangle \longrightarrow |a\rangle \otimes \frac{1}{\sqrt{Q}} \sum_{b=0}^{Q-1} |b\rangle \longrightarrow |a\rangle \otimes \frac{1}{\sqrt{Q}} \sum_{b=0}^{Q-1} w^{ab} |b\rangle = |a\rangle \otimes |\chi_a\rangle$$

We can implement the circuit that sends $|0\rangle \longrightarrow \frac{1}{\sqrt{Q}} \sum_{b=0}^{Q-1} |b\rangle$ efficiently in the following two ways:

1. Perform the following sequence of transformations.

$$|0\rangle^m \otimes |0\rangle \stackrel{H^{\otimes m}}{\longrightarrow} \frac{1}{M} \sum_{x=0}^{2^m-1} |x\rangle|0\rangle \stackrel{x \geq Q}{\longrightarrow} \frac{1}{M} \sum_{x=0}^{2^m-1} |x\rangle|x \geq Q\rangle$$

   Note that since we can efficiently decide whether or not $x \geq Q$ classically, we can also do so quantum mechanically. Now take measurement on the second register. If the result is a 0, the first register contains a uniform superposition over $|0\rangle, \ldots, |Q-1\rangle$. If not, we repeat the experiment. At each trial, we succeed with probability $Q/M > 2^{m-1}/2^m = 1/2$.

2. If we pick a number u.a.r. in the range 0 to $Q-1$, the most significant bit of the number is 0 with probability $2^{m-1}/Q$. We can therefore set the first bit of our output to be the superposition:

$$\sqrt{\frac{2^{m-1}}{Q}}|0\rangle + \sqrt{1 - \frac{2^{m-1}}{Q}}|1\rangle$$

If the first bit is 0, then the remaining $m-1$ bits may be chosen randomly and independently, which correspond to the output of $H^{\otimes m-1}$ on $\left|0^{m-1}\right\rangle$. If the first bit is 1, we need to pick the remaining $m-1$ bits to correspond to a uniformly chosen random number between 0 and $Q-2^{m-1}$, which we can do recursively.

The second transformation $\left|a\right\rangle\left|b\right\rangle \rightarrow w^{ab}\left|a\right\rangle\left|b\right\rangle$ can be made using the controlled phase shift circuit.

This gives us an efficient quantum circuit for $\left|a\right\rangle\left|0\right\rangle \rightarrow \left|a\right\rangle\left|\chi_a\right\rangle$, but what we really want is a circuit for $\left|a\right\rangle \rightarrow \left|\chi_a\right\rangle$. In particular, for application to factoring, we need a circuit that "forgets" the input a in order to have interference in the superposition over $\left|\chi_a\right\rangle$.

What we would like is a quantum circuit that transforms $\left|a\right\rangle\left|\chi_a\right\rangle \rightarrow \left|0\right\rangle\left|\chi_a\right\rangle$. If we could find a unitary transformation U with eigenvector $\left|\chi_a\right\rangle$ and eigenvalue $e^{2\pi i a/Q}$, then we could use phase estimation to implement the transformation $\left|0\right\rangle\left|\chi_a\right\rangle \rightarrow \left|a\right\rangle\left|\chi_a\right\rangle$. By reversing the quantum circuit for phase estimation (which we could do since quantum circuits are reversible), we have an efficient quantum circuit for

$$\left|a\right\rangle\left|0\right\rangle \rightarrow \left|a\right\rangle\left|\chi_a\right\rangle \rightarrow \left|0\right\rangle\left|\chi_a\right\rangle$$

which is what we need. Note that the phase estimation circuit with m bits of precision outputs j such that $\frac{j}{2^m} \approx \frac{a}{Q}$. So if we take $2^m >> Q^2$, we can use continued fractions to reconstruct a as required above.

To see that the required U exists, consider $U : \left|x\right\rangle \rightarrow \left|x-1 \bmod Q\right\rangle$. Then,

$$U(\chi_a) = U\left(\sum_{b=0}^{Q-1} w^{ab}\left|b\right\rangle\right) = \sum_{b=0}^{Q-1} w^{ab}\left|b-1\right\rangle = w^a \sum_{b=1}^{Q} w^{a(b-1)}\left|b-1\right\rangle = w^a \chi_a.$$

In addition, note that U^k can be efficiently computed with a classical circuit, and can therefore be both efficiently and reversibly computed with a quantum circuit. The overall circuit to compute QFT mod Q is shown in Figure **??** (The circuit should be read from right to left).

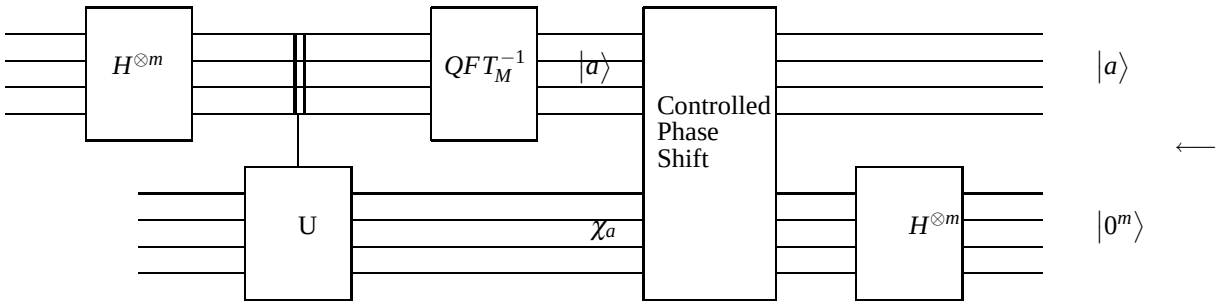

Figure 0.7: Using Reverse Phase Estimation Circuit to do QFT mod Q for arbitrary Q

0.4 Mixed Quantum State

Next, we will outline a very interesting application of phase estimation. Before we can do this, we must introduce some concepts in quantum information theory.

So far we have dealt with *pure* quantum states

$$|\psi\rangle = \sum_x \alpha_x |x\rangle.$$

This is not the most general state we can think of. We can consider a probability distribution of pure states, such as $|0\rangle$ with probability $1/2$ and $|1\rangle$ with probability $1/2$. Another possibility is the state

$$\begin{cases} |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) & \text{with probability 1/2} \\ |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) & \text{with probability 1/2} \end{cases}$$

In fact, no measurement can distinguish the first case $\left(|0\rangle \text{ or } |1\rangle\right)$ from this case. This will be seen below.

In general, we can think of *mixed* state $\{p_i, |\psi_i\rangle\}$ as a collection of pure states $|\psi_i\rangle$, each with associated probability p_i, with the conditions $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$. One context in which mixed states arise naturally is in quantum protocols, where two players share an entangled (pure) quantum state. Each player's view of their quantum register is then a probability distribution over pure states (achieved when the other player measures their register). Another reason we consider such mixed states is because the quantum states are hard to isolate, and hence often entangled to the environment.

0.5 Density Matrix

Now we consider the result of measuring a mixed quantum state. Suppose we have a mixture of quantum states $|\psi_i\rangle$ with probability p_i. Each $|\psi_i\rangle$ can be represented by a vector in $\mathscr{C}^{2^n}$, and thus we can associate the outer product $|\psi_i\rangle\langle\psi_i| = \psi_i\psi_i^*$, which is an $2^n \times 2^n$ matrix

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} \begin{pmatrix} \bar{a_1} & \bar{a_2} & \cdots & \bar{a_N} \end{pmatrix} = \begin{pmatrix} a_1\bar{a_1} & a_1\bar{a_2} & \cdots & a_1\bar{a_N} \\ a_2\bar{a_2} & a_1\bar{a_2} & \cdots & a_2\bar{a_N} \\ \vdots & & & \vdots \\ a_N\bar{a_1} & a_N\bar{a_2} & \cdots & a_N\bar{a_N} \end{pmatrix}.$$

We can now take the average of these matrices, and obtain the *density matrix* of the mixture $\{p_i, |\psi_i\rangle\}$:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|.$$

We give some examples. Consider the mixed state $|0\rangle$ with probability of $1/2$ and $|1\rangle$ with probablity $1/2$. Then

$$|0\rangle\langle 0| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$|1\rangle\langle 1| = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus in this case

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Now consider another mixed state, this time consisting of $|+\rangle$ with probability $1/2$ and $|-\rangle$ with probability $1/2$. This time we have

$$|+\rangle\langle +| = (1/2)\begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

and

$$|-\rangle\langle -| = (1/2)\begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Thus in this case the offdiagonals cancel, and we get

$$\rho = \frac{1}{2}|+\rangle\langle+| + \frac{1}{2}|-\rangle\langle-| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Note that the two density matrices we computed are identical, even though the mixed state we started out was different. Hence we see that it is possible for two different mixed states to have the same density matrix.

Nonetheless, the density matrix of a mixture completely determines the effects of making a measurement on the system:

Theorem 0.1: *Suppose we measure a mixed state $\{p_j, |\psi_j\rangle\}$ in an orthonormal bases $|\beta_k\rangle$. Then the outcome is $|\beta_k\rangle$ with probability $\langle\beta_k|\rho|\beta_k\rangle$.*

Proof: We denote the probability of measuring $|\beta_k\rangle$ by $\Pr[k]$. Then

$$\begin{aligned} \Pr[k] &= \sum_j p_j |\langle\psi_j|\beta_k\rangle|^2 \\ &= \sum_j p_j \langle\beta_k|\psi_j\rangle\langle\psi_j|\beta_k\rangle \\ &= \left\langle \beta_k \middle| \sum_j p_j |\psi_j\rangle\langle\psi_j| \middle| \beta_k \right\rangle \\ &= \langle\beta_k|\rho|\beta_k\rangle. \end{aligned}$$

□

Thus mixtures with the same density matrix are indistinguishable by measurement. It will be shown in the next section that, in fact, two mixtures are distinguishable by measurement if and only if they have different density matrices.

We list several properties of the density matrix:

1. ρ is Hermitian, so the eigenvalues are real and the eigenvectors orthogonal.
2. If we measure in the standard basis the probability we measure i, $P[i] = \rho_{i,i}$. Also, the eigenvalues of ρ are non-negative. Suppose that λ and $|e\rangle$ are corresponding eigenvalue and eigenvector. Then if we measure in the eigenbasis, we have
$$\Pr[e] = \langle e|\rho|e\rangle = \lambda\langle e|e\rangle = \lambda.$$
3. $\mathrm{tr}\rho = 1$. This is because if we measure in the standard basis $\rho_{i,i} = Pr[i]$ but also $\sum_i Pr[i] = 1$ so that $\sum_i \rho_{i,i} = \sum_i Pr[i] = 1$.

Consider the following two mixtures and their density matrices:

$$\left.\begin{array}{llll} \cos\theta|0\rangle + \sin\theta|1\rangle & \text{w.p. } 1/2 & = \frac{1}{2}\begin{pmatrix} c\theta \\ s\theta \end{pmatrix}\begin{pmatrix} c\theta & s\theta \end{pmatrix} & = \frac{1}{2}\begin{pmatrix} c^2\theta & c\theta s\theta \\ c\theta s\theta & s^2\theta \end{pmatrix} \\ \cos\theta|0\rangle - \sin\theta|1\rangle & \text{w.p. } 1/2 & = \frac{1}{2}\begin{pmatrix} c\theta \\ -s\theta \end{pmatrix}\begin{pmatrix} c\theta & -s\theta \end{pmatrix} & = \frac{1}{2}\begin{pmatrix} c^2\theta & -c\theta s\theta \\ -c\theta s\theta & s^2\theta \end{pmatrix} \end{array}\right\} = \begin{pmatrix} \cos^2\theta & 0 \\ 0 & \sin^2\theta \end{pmatrix}$$

$$\left.\begin{array}{llll} |0\rangle & \text{w.p.} \cos^2\theta & = \cos^2\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix} & = \cos^2\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ |1\rangle & \text{w.p. } \sin^2\theta & = \sin^2\theta\begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix} & = \sin^2\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{array}\right\} = \begin{pmatrix} \cos^2\theta & 0 \\ 0 & \sin^2\theta \end{pmatrix}$$

Thus, since the mixtures have identical density matrices, they are indistinguishable.

0.6 Von Neumann Entropy

We will now show that if two mixed states are represented by different density measurements, then there is a measurement that distinguishes them. Suppose we have two mixed states, with density matrices A and B such that $A \neq B$. We can ask, what is a good measurement to distinguish the two states? We can diagonalize the difference $A-B$ to get $A-B = E\Lambda E^*$, where E is the matrix of orthogonal eigenvectors. Then if e_i is an eigenvector with eigenvalue λ_i, then λ_i is the difference in the probability of measuring e_i:

$$\Pr_A[i] - \Pr_B[i] = \lambda_i.$$

We can define the distance between two probability distributions (with respect to a basis E) as

$$|\mathscr{D}_A - \mathscr{D}_B|_E = \sum \left(\Pr_A[i] - \Pr_B[i]\right).$$

If E is the eigenbasis, then

$$|\mathscr{D}_A - \mathscr{D}_B|_E = \sum_i |\lambda_i| = \mathrm{tr}|A-B| = \|A-B\|_{\mathrm{tr}},$$

which is called the trace distance between A and B.

Claim Measuring with respect to the eigenbasis E (of the matrix $A-B$) is optimal in the sense that it maximizes the distance $|\mathscr{D}_A - \mathscr{D}_B|_E$ between the two probability distributions.

Before we prove this claim, we introduce the following definition and lemma without proof.

Definition Let $\{a_i\}_{i=1}^N$ and $\{b_i\}_{i=1}^N$ be two non-increasing sequences such that $\sum_i a_i = \sum_i b_i$. Then the sequence $\{a_i\}$ is said to majorize $\{b_i\}$ if for all k,

$$\sum_{i=1}^{k} a_i \geq \sum_{i=1}^{k} b_i.$$

Lemma[Schur] Eigenvalues of any Hermitian matrix majorizes the diagonal entries (if both are sorted in nonincreasing order).

Now we can prove our claim.

Proof Since we can reorder the eigenvectors, we can assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Note that $\mathrm{tr}(A-B) = 0$, so we must have $\sum_i \lambda_i = 0$. We can split the λ_i's into two groups: positive ones and negative ones, we must have

$$\sum_{\lambda_i > 0} = \frac{1}{2}\|A-B\|_{\mathrm{tr}} \qquad \sum_{\lambda_i < 0} = -\frac{1}{2}\|A-B\|_{\mathrm{tr}}.$$

Thus

$$\max_k \sum_{i=1}^{k} \lambda_i = \frac{1}{2}\|A-B\|_{\mathrm{tr}}.$$

Now consider measuring in another basis. Then the matrix $A-B$ is represented as $H = F(A-B)F^*$, and let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ be the diagonal entries of H. Similar argument shows that

$$\max_k \sum_{i=1}^{k} \mu_i = \frac{1}{2}\sum_{i=1}^{n} |\mu_i| = \frac{|\mathscr{D}_A - \mathscr{D}_B|_F}{2}.$$

But by Schur's lemma the λ_i's majorizes μ_i's, so we must have

$$|\mathscr{D}_A - \mathscr{D}_B|_F \leq |\mathscr{D}_A - \mathscr{D}_B|_E = \|A-B\|_{\mathrm{tr}}.$$

Let $H(X)$ be the *Shannon Entropy* of a random variable X which can take on states $p_1 \dots p_n$.

$$H(\{p_i\}) = \sum_i p_i \log \frac{1}{p_i}$$

In the quantum world, we define an analogous quantity, $S(\rho)$, the *Von Neumann entropy* of a quantum ensemble with density matrix ρ with eigenvalues $\lambda_1, \dots, \lambda_n$:

$$S(\rho) = H\{\lambda_1, \dots, \lambda_n\} = \sum_i \lambda_i \log \frac{1}{\lambda_i}$$

0.7 Phase Estimation and Mixed State Computation

Liquid NMR (Nuclear Magnetic Resonance) quantum computers have successfully implemented 7 qubits and performed a stripped down version of quantum factoring on the number 15. In liquid NMR, the quantum register is composed of the nuclear spins in a suitably chosen molecule - the number of qubits is equal to the number of atoms in the molecule. We can think of the computer as consisting of about 10^{16} such molecules (a macroscopic amount of liquid), each controlled by the same operations simultaneously. Thus we will have 10^{16} copies of our state, each consisting of say 7 qubits. We assume that we can address the qubits individually, so that for example, we could preform an operation such as *CNOT* on the 2nd and 4th qubit (simultaneously on each copy).

The catch in liquid NMR quantum computing is that initializing the register is hard. Each qubit starts out in state $|0\rangle$ with probability $1/2+\varepsilon$ and in state $|1\rangle$ with probability $1/2-\varepsilon$. Here ε depends upon the strength of the magnetic field that the liquid sample is placed in. Using very strong magnets in the NMR apparatus, the polarization ε is still about 10^{-5}.

If $\varepsilon = 0$ then the density matrix describing the quantum state of the register is $\rho = \frac{1}{2^n} I$. This means that if we apply a unitary transformation U, the density matrix of the resulting state is $I \rightarrow_U UIU^\dagger = I$. So you cannot perform any meaningful computation.

The way NMR quantum computation works is this: the initial mixed state (with $\varepsilon = 10^{-5}$) is preprocessed (through a sequence of quantum gates) to obtain a new mixed state which is maximally mixed ($\frac{1}{2^n} I$) with probability $1-\delta$ and $|0000000\rangle$ with probability δ. Now, if we apply a unitary transformation to this state, we get $\frac{1}{2^n} I$ with probability $1-\delta$ and $U|0000000\rangle$ with probability δ. Thus if we measure the state, we obtain a coin flip with bias δ^2 towards the correct answer. Another way of thinking about this is that the $\frac{1}{2^n} I$ gives no net signal in the measurement, while the δ^2 signal gets amplified by the 10^{16} copies of the computation being carried out simultaneously. The problem is that δ is exponentially small in n the number of qubits. Therefore liquid NMR quantum computation cannot scale beyond 10-20 qubits.

What if we tried one qubit with ε-bias and n-1 qubits maximally mixed?

Question: Say we have a single clean bit (and n maximally mixed qubits), what can we do with this?

We can do at least one quantum computation, phase estimation to approximate the trace of a unitary matrix. Use the single clean qubit as the control bit and apply the (controlled) unitary to the n maximally mixed qubits. We can think of the n qubits as being a uniform mixture over the eigenvectors of the unitary, and upon measuring, we get out a random eigenvalue estimate for trace(U). See Figure **??**

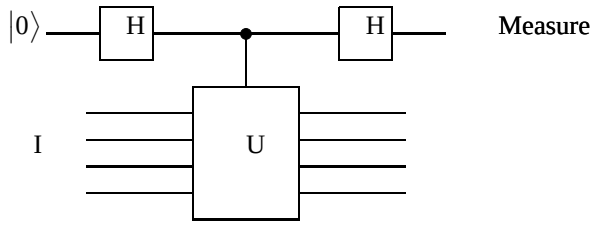

Figure 0.8: Phase Estimation Circuit for Trace(U)

Is there anything else that we can do with just one qubit? Can you prove limits on what can be done with one clean qubit? And where is the entanglement in the computation?