Quantum Computing

Muchang Bahng

Spring 2024

Contents

1 Fundamentals

1.1 Qubits

Definition 1.1 (Qubit)

The simplest quantum mechanical system is the qubit, which is described by a wavefunction defined on the position space $X = \{0, 1\}$. Any complex-valued function ψ on X must be a linear combination of two the two delta functions δ_0 and δ_1, and so we can write the general wavefunction as

$$\psi = \alpha\delta_0 + \beta\delta_1 = \alpha 0 + \beta 1 \tag{1}$$

where $0, 1$ are not scalars but notation for the delta functions. The coefficients of $|\psi\rangle$ are called the amplitudes.

The ket notation unifies this set of function and the elements in the domain X into a single complex Hilbert space $\mathcal{H}$, and so we can write this equivalently in ket notation as

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{2}$$

The space $L^2(\{0, 1\})$ in which ψ lives in is a complex 2-dimensional vector space, and so we can write its elements as column vectors with each element representing the coefficients or the values of the wavefunction at each point.

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, |\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \tag{3}$$

Quantum computing is really just quantum mechanics, but with the main difference that now we are looking at discrete X, and therefore a finite-dimensional $\mathcal{H}$. Generalization of all the rules that we need are covered in quantum mechanics, so we will briefly mention them here.

Definition 1.2 (Basis States)

We list 4 different orthonormal bases that we will work with often in $\mathbb{C}^2$. We first introduce the standard basis, the Z basis state, and write the rest of the states in the Z basis.

1. The classical notions of the 0 and 1 bit can be represented as the orthonormal vectors $|0\rangle, |1\rangle \in \mathbb{C}^2$, called **computational basis states** or the **z basis states**, which are

$$|0\rangle = |+\rangle_z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = |-\rangle_z = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

2. The **x basis states** are

$$|+\rangle_x := \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \quad |-\rangle_x := \frac{|0\rangle - |1\rangle}{\sqrt{2}} \tag{4}$$

3. The **y basis states** are

$$|+\rangle_y := \frac{|0\rangle + i\,|1\rangle}{\sqrt{2}}, \quad |-\rangle_y := \frac{|0\rangle - i\,|1\rangle}{\sqrt{2}} \tag{5}$$

Example 1.1 ()

Say that a qubit with respect to the standard basis can be written as

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}_Z \tag{6}$$

where the subscript Z implies that we are in the Z basis. Then, we can write it with respect to another basis as

$$|\psi\rangle = \alpha\frac{|+\rangle + |-\rangle}{\sqrt{2}} + \beta\frac{|+\rangle - |-\rangle}{\sqrt{2}} = \frac{\alpha+\beta}{\sqrt{2}}|+\rangle + \frac{\alpha-\beta}{\sqrt{2}}|-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}\alpha+\beta\\ \alpha-\beta\end{pmatrix}_X \tag{7}$$

Now, we can measure this same qubit with a new observable with respect to this new basis.

Example 1.2 ()

If we take $|\psi\rangle$ in the X-basis representation and measure it with the position operator (also in the X-basis), which we represent

$$\hat{M} = \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}_x \tag{8}$$

Then its eigenvectors are

$$|+\rangle = \begin{pmatrix}1\\0\end{pmatrix}_x = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}_z \quad |-\rangle = \begin{pmatrix}0\\1\end{pmatrix}_x = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}_z \tag{9}$$

and so we can decompose it into

$$|\psi\rangle = \frac{\alpha+\beta}{\sqrt{2}}|+\rangle + \frac{\alpha-\beta}{\sqrt{2}}|-\rangle \implies \begin{cases}\mathbb{P}(M=0) = \frac{|\alpha+\beta|^2}{2}\\ \mathbb{P}(M=1) = \frac{|\alpha-\beta|^2}{2}\end{cases} \tag{10}$$

We can do this entire process equivalently in the Z-basis as well. Note that

$$\hat{M} = \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}_x = \begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}_z \tag{11}$$

with eigenvectors $|+\rangle$ (with eigenvalue 0) and $|-\rangle$ (eigenvalue 1), and so we can again decompose the original Z-expansion into the X-expansion and see that the coefficients are the same.

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \frac{\alpha+\beta}{\sqrt{2}}|+\rangle + \frac{\alpha-\beta}{\sqrt{2}}|-\rangle \implies \begin{cases}\mathbb{P}(M=0) = \frac{|\alpha+\beta|^2}{2}\\ \mathbb{P}(M=1) = \frac{|\alpha-\beta|^2}{2}\end{cases} \tag{12}$$

Therefore, once we pick any orthonormal basis, we can always express its state in that basis and measure it in that basis as well, and we even have the flexibility to use basis transformations to measure things in a basis that is not what the state is expressed in.

1.2 Measurements

1.2.1 Projective Measurements

The act of measuring a quantum system with a Hermitian observable is known as a **projective measurement**. To review measurements and for ease of computation, we provide an equivalent way to compute the probabilities of the outcomes of a measurement.

1. We start off with a Hermitian operator $\hat{M}$, which can be decomposed into a linear combination of mutually orthogonal projection operators along with their respective eigenvalues.

$$\hat{M} = \sum_m \lambda_m M_m = \sum_m \lambda_m |q_m\rangle\langle q_m| \tag{13}$$

2. The state vector $|\psi\rangle$ can also be expanded into the eigenbasis of $\hat{M}$ and we can get its respective

coefficients of the ith eigenvalue by

$$|\psi\rangle = \sum_m \alpha_m |q_m\rangle \implies \langle q_m|\psi\rangle = \alpha_m \tag{14}$$

3. Therefore, the probability that M will realize onto the mth eigenvalue is the modulus squared of the coefficient of the mth eigenvalue.

$$\mathbb{P}(M = \lambda_m) = |\langle q_m|\psi\rangle|^2 = \langle\psi| P_m^\dagger P_m |\psi\rangle = \langle\psi| P_m |\psi\rangle = |\alpha_m|^2 \tag{15}$$

4. The state of the system after the measurement is the (normalized) projection of the state vector onto the eigenspace of the realized eigenvalue.

$$|\psi\rangle \mapsto \frac{P_m|\psi\rangle}{||P_m|\psi\rangle||} = \frac{|q_m\rangle\langle q_m|\psi\rangle}{||\langle q_m|\psi\rangle||} = \frac{|q_m\rangle\langle q_m|\psi\rangle}{||\alpha_m||} = \frac{|q_m\rangle\langle q_m|\psi\rangle}{\sqrt{\mathbb{P}(M=\lambda_m)}} \tag{16}$$

Definition 1.3 (Expection, Variance of Observable)

The expectation of M is

$$\begin{aligned}\mathbb{E}(M) &= \sum_m m\, p(m) \\ &= \sum_m m\langle\psi|P_m|\psi\rangle \\ &= \langle\psi|\left(\sum_m mP_m\right)|\psi\rangle \\ &= \langle\psi|M|\psi\rangle \equiv \langle M\rangle\end{aligned}$$

with variance

$$\begin{aligned}\big(\Delta M\big)^2 &= \langle (M - \langle M\rangle)^2\rangle \\ &= \langle M^2\rangle - \langle M\rangle^2\end{aligned}$$

Example 1.3 (Basic Measurement of a Qubit)

Let us have a 1-qubit quantum system that is in the state

$$|\psi\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} = \begin{pmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix}$$

and we (projectively) observe it with the Pauli-Z operator. We can calculate it to have eigenvalue $+1$ with eigenvector $|0\rangle$ and eigenvalue -1 with eigenvector $|1\rangle$. The decomposition of Z into its projective maps is

$$\begin{aligned}Z &= (+1)\, P_{+1} + (-1)\, P_{-1} \\ &= (+1)\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} + (-1)\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\end{aligned}$$

and so the probability of getting a measurement of $+1$ or -1 is

$$p(+1) = \langle\psi|P_{+1}|\psi\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{2}$$

$$p(-1) = \langle\psi|P_{-1}|\psi\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{2}$$

Note that so far, we've talked about measurement operators that are Hermitian. This automatically gives us an orthonormal basis of eigenvectors, but the eigendecomposition may not be a set of 1-dimensional eigenspaces. For example, the identity operator has one eigenvalue of 1 with a 2-dimensional eigenspace. This is called a partial projective measurement (since we're not projecting it "fully" onto a 1-dimensional eigenspace). Therefore, we must distinguish between these two.

Definition 1.4 (Complete Projective Measurement)

A **complete projective measurement** is a projective measurement with a Hermitian operator that has eigenspaces of dimension 1.

Definition 1.5 (Partial Projective Measurement)

A **partial projective measurement** is a projective measurement with a Hermitian operator that has eigenspaces of dimension greater than 1.

If a partial projective measurement $\hat{M}$ has eigenvalue λ with a d-dimensional eigenspace, then observing the value $M = \lambda$ will only project the state vector $|\psi\rangle$ into that eigenspace. This can be very useful since it still keeps it somewhat in superposition, retaining some of the original information in the original state $|\psi\rangle$. A classic application is the parity operator, which we will see later.

1.2.2 General Measurements

Theorem 1.1 (Posulate 3: General Measurement)

Given a state vector $|\psi\rangle \in \mathcal{H}$ with a total possible number of measurement outcomes parameterized by m. Then, a quantum measurement is described by a collection $\{M_m\}$ of **measurement operators** (acting on the state space) satisfying the *completeness equation*

$$\sum_m M_m^\dagger M_m = I$$

If the state of the quantum system is $|\psi\rangle$ immediately before the measurement, then the probability that result m occurs is

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle$$

Immediately after this measurement outcome m, the state of the system then becomes

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}} = \frac{M_m|\psi\rangle}{\sqrt{p(m)}}$$

Note that the completeness equation implies

$$1 = \sum_m p(m) = \sum_m \langle\psi|M_m^\dagger M_m|\psi\rangle$$

Note that we have defined measurements in a basis-independent way. If we do fix an orthonormal $B = \{|\psi_1\rangle, \dots, |\psi_n\rangle\}$, then both the state $|\psi\rangle$ and the measurement operators $\{M_m\}$ should be written with respect to B.

That's all there is to measurements: they are a collection of operators satisfying the normalization identity above. Usually, this is not how measurement operators are introduced in quantum mechanics courses. They are introduced as projective measurements which can be encoded in self-adjoint operators, but for more precise measurements needed in quantum computing, we should introduce the generalized version first.

Example 1.4 (Simple Example of General Measurement)

As an example, suppose that we have the simple quantum system consisting of a single qubit that has the state $|\psi\rangle = a|0\rangle + b|1\rangle$ (w.r.t. Z basis) immediately before measurement

$$\{M_0, M_1\} = \left\{ \begin{pmatrix} 0.8 & 0 \\ 0 & 0.6 \end{pmatrix}, \begin{pmatrix} 0.6 & 0 \\ 0 & 0.8 \end{pmatrix} \right\}$$

which clearly satisfies the completeness equation $M_0^\dagger M_0 + M_1^\dagger M_1 = I$, has probabilities

1. $p(0) = \langle\psi|M_0^\dagger M_0|\psi\rangle = 0.64a^2 + 0.36b^2$ chance of a measurement outcome of 0
2. $p(1) = \langle\psi|M_1^\dagger M_1|\psi\rangle = 0.36a^2 + 0.64b^2$ chance of a measurement outcome of 1

If we observe a measurement of 0 in the system, then the state vector of the quantum system would be

$$\frac{M_0|\psi\rangle}{\sqrt{\langle\psi|M_0^\dagger M_0|\psi\rangle}} = \frac{1}{\sqrt{0.64a^2 + 0.36b^2}} \begin{pmatrix} 0.8a \\ 0.6b \end{pmatrix} = \frac{0.8a}{\sqrt{0.64a^2 + 0.36b^2}} |0\rangle + \frac{0.6b}{\sqrt{0.64a^2 + 0.36b^2}} |1\rangle$$

and if we observe an outcome of 1, then the state vector of the system would be

$$\frac{M_1|\psi\rangle}{\sqrt{\langle\psi|M_1^\dagger M_1|\psi\rangle}} = \frac{1}{\sqrt{0.36a^2 + 0.64b^2}} \begin{pmatrix} 0.6a \\ 0.8b \end{pmatrix} = \frac{0.6a}{\sqrt{0.36a^2 + 0.64b^2}} |0\rangle + \frac{0.8b}{\sqrt{0.36a^2 + 0.64b^2}} |1\rangle$$

Theorem 1.2 (Composition of Measurements)

A following theorem is that a composition of measurements, e.g., $\{L_l\}$ followed by a separate $\{M_m\}$ is physically equivalent to a single measurement defined by measurement operators $\{N_{lm}\}$ with the representation $N_{lm} = M_m L_l$.

1.2.3 POVM Measurements

The measurement postulate involves two elements: the probabilities of the measurement outcomes and the post-measurement state of the system. When we are concerned with only the probabilities (e.g., in the case of an experiment where the system is measured only once), it is useful to employ the POVM formalism. Suppose a measurement described by measurement operators M_m is performed upon a quantum system in the state $|\psi\rangle$. Then, the probability of outcome m is given by $p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle$.

We now define

$$E_m \coloneqq M_m^\dagger M_m$$

to be the **POVM elements** associated with the measurement, and the complete set $\{E_m\}$ to be the **POVM**. Some linear algebra reveals that E_m must be positive-definite, the POVM is sufficient to completely determine the probabilities of the different measurement outcomes. For a projective measurement described by measurement operators P_m, we can see that they are equal to the POVM elements since $E_m \equiv P_m^\dagger P_m = P_m$.

We have seen that the existence of a set of measurement operators $\{M_m\}$ satisfying the completeness equation automatically implies the existence of the POVM $\{E_M\}$ consisting of positive operators satisfying $\sum_m E_m = I$:

$$\{M_m\} \longrightarrow \{E_m\}.$$

Going backwards, we claim that the existence of an arbitrary set of positive operators $\{E_M\}$ satisfying $\sum_m E_m = I$ implies the existence of measurement operators $\{M_m\}$ defining a measurement described by the POVM. We simply define $M_m \equiv \sqrt{E_m}$, which we can do since E_m is positive (define a new linear map with the same eigenspaces but square root of eigenvalues).

$$\{M_m\} \longleftrightarrow \{E_m\}.$$

The applicability of POVMs is demonstrated in the following example: Suppose a qubit is in one of two states: $|\psi_1\rangle = |0\rangle$ or $|\psi_2\rangle = (|0\rangle + |1\rangle)/2$. Since these qubits are not orthonormal, we cannot determine the state with 100% accuracy. However, it is possible for us to perform a measurement that distinguishes the states sometimes, but never makes an error of identification. We can construct a POVM of three elements as such:

$$\begin{aligned}
E_1 &\equiv \frac{\sqrt{2}}{1+\sqrt{2}}|1\rangle\langle 1| = \frac{\sqrt{2}}{1+\sqrt{2}}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \\
E_2 &\equiv \frac{\sqrt{2}}{2+2\sqrt{2}}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \\
E_3 &\equiv I - E_1 - E_2.
\end{aligned}$$

It can be checked that these sum up to I and are positive definite. If the actual state of the system was $|\psi_1\rangle$, then we have

$$\begin{aligned}
p(1 \,|\, |\psi_1\rangle) &= \langle\psi_1|E_1|\psi_1\rangle = 0, \\
p(2 \,|\, |\psi_1\rangle) &= \langle\psi_1|E_2|\psi_1\rangle = \frac{\sqrt{2}}{2+2\sqrt{2}}, \\
p(3 \,|\, |\psi_1\rangle) &= \langle\psi_1|E_3|\psi_1\rangle = \frac{2+\sqrt{2}}{2+2\sqrt{2}}.
\end{aligned}$$

And if the actual state of the system was $|\psi_2\rangle$, then we have

$$\begin{aligned}
p(1 \,|\, |\psi_2\rangle) &= \langle\psi_2|E_1|\psi_2\rangle = \frac{\sqrt{2}}{2+2\sqrt{2}}, \\
p(2 \,|\, |\psi_2\rangle) &= \langle\psi_2|E_2|\psi_2\rangle = 0, \\
p(3 \,|\, |\psi_2\rangle) &= \langle\psi_2|E_3|\psi_2\rangle = \frac{2+\sqrt{2}}{2+2\sqrt{2}}.
\end{aligned}$$

Clearly, we can see that if the measurement outcome yields 1, then the actual state of the system must have been $|\psi_2\rangle$, and if it yields 2, then the actual state must have been $|\psi_1\rangle$. In the case where the outcome is 3, then we would not know, but at least there is no risk of misinterpreting.

1.3 Phase Factors and the Bloch Sphere

Definition 1.6 (Global and Relative Phase)

Note that the unit sphere in $S^2 \subset \mathbb{C}^2$ captures the state space of the single qubit in full generality. Because $|\alpha|^2 + |\beta|^2 = 1$, we can rewrite the qubit to

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle = e^{i\gamma}\left(\cos\frac{\theta}{2}\,|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}\,|1\rangle\right) \tag{17}$$

where $\theta, \gamma, \varphi \in \mathbb{R}$. The $e^{i\gamma}$ is known as the **global phase** and $e^{i\varphi}$ is known as the **relative phase**.

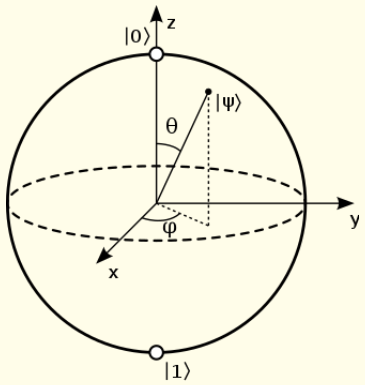
Now note that $|\psi\rangle$ and $e^{-i\gamma}\,|\psi\rangle$ are two different states in S^2. Let us measure them with a set of measurements $\{M_m\}$, and we can see that they both give the same probabilities for each outcome m.

$$\begin{aligned} p(m) &= \langle e^{i\delta}\psi | M_m^\dagger M_m | e^{i\delta}\psi\rangle, \\ &= \langle \psi | e^{-i\delta} M_m^\dagger M_m e^{i\delta} | \psi\rangle, \\ &= \langle \psi | M_m^\dagger M_m | \psi\rangle = p(m) \end{aligned}$$

Therefore, both $|\psi\rangle$ and any state vector of form $e^{i\gamma}\,|\psi\rangle$ produce the same probabilities under any general measurement, and from an observational point of view, these two states are identical. So we can construct a quotient space on S^2 by defining a equivalence relation $|\psi\rangle \sim e^{i\gamma}\,|\psi\rangle$ where both are equal up to a global phase factor. This restricts our states space to having 2 real parameters θ and φ, so we now can construct some visual.

Definition 1.7 (Bloch Sphere)

The previous parameterization of the real unit sphere $S^2 \subset \mathbb{R}^2$ is known as the **Bloch sphere**, which can be visualized below.



A few properties should be mentioned:

1. The antiparallel vectors lying on the X, Y, Z axes represent the orthonormal X, Y, Z basis in $\mathbb{C}^2$. Note that while the basis is pairwise perpendicular in $\mathbb{C}^2$, in the Bloch sphere they are visualized as antiparallel.
2. The probability of it being $|0\rangle$ or $|1\rangle$ depends on the value of θ.

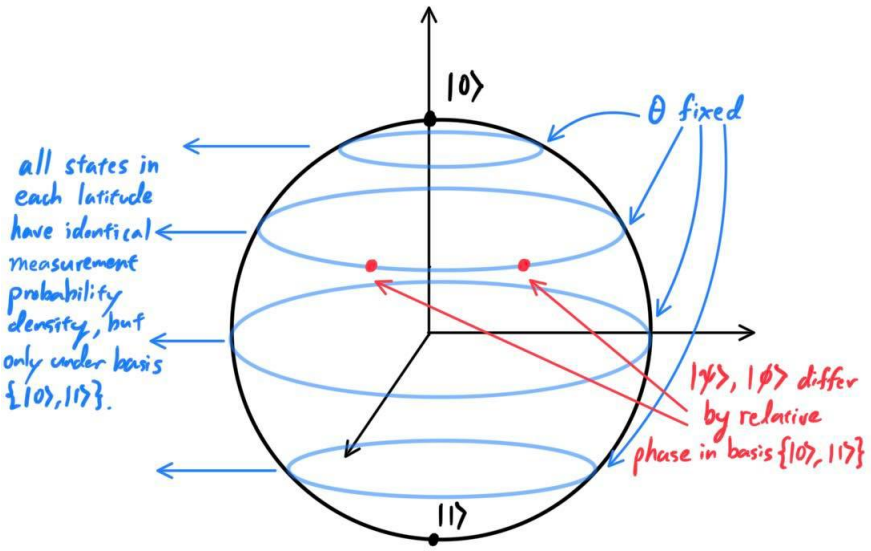
The other kind of phase is known as the **relative phase factor**. Given two states

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \text{ and } |\psi^*\rangle = \alpha^*|0\rangle + |\beta^*\rangle,$$

if $|\alpha| = |\alpha^*|$ or $|\beta| = |\beta^*|$, then we say that **the amplitudes differ by a relative phase**. Furthermore, two states $|\psi\rangle, |\psi^*\rangle$ are said to **differ by a relative phase in some basis** if each of the amplitudes in that basis is related by such a phase factor. For example, two states

$$\frac{|0\rangle + |1\rangle}{\sqrt{2}} \text{ and } \frac{|0\rangle - |1\rangle}{\sqrt{2}}$$

differ by a relative phase (in the computational basis $|0\rangle, |1\rangle$) since the $|0\rangle$ amplitudes differ by a relative phase factor of 1 ($\frac{1}{\sqrt{2}} = 1 \cdot \frac{1}{\sqrt{2}}$) and the $|1\rangle$ amplitudes differ by a relative phase factor of -1 ($-\frac{1}{\sqrt{2}} = -1 \cdot \frac{1}{\sqrt{2}}$). It is clear that due to Born's rule on this one-qubit system, $|\alpha| = |\alpha^*| \iff |\beta| = |\beta^*|$, and so, all we have to do is check the magnitudes of the $|0\rangle$ amplitudes of two state vectors. Visualizing this on the Bloch sphere, we can see that the θ is the only parameter capable of changing the $|0\rangle$ amplitude. The global phase factor $e^{i\gamma}$ is merely a rotation map and also cannot change the $|0\rangle$. Therefore, we can see that two state vectors differ by a relative phase if and only if they have the same θ value, i.e. if the two points on the Bloch sphere are on the same "latitude."



Notice that if two states are differ by a relative phase, then these phases are observationally equivalent, and so must be similar to the global phase factor. However, the relative phase is basis-dependent and so may produce different probability densities depending on the computational basis, while the global one is basis-independent.

1.4 Entanglement

If we are interested in a composite quantum system made up of two (or more) distinct physical systems, the states of the composite system can be described as stated in postulate 4.

Theorem 1.3 (Postulate 4: Composite Systems)

The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. That is, if we have systems $1, \dots, n$ with the state vector of the ith system being $|\psi_i\rangle$, then the joint state of the total system is

$$\bigotimes_i |\psi_i\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \dots \otimes |\psi_n\rangle.$$

A further property of the tensor product of Hilbert spaces is the induced inner product. That is, if H_1 and H_2 are Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, respectively, then $H_1 \otimes H_2$ is a Hilbert space with an induced inner product

$$\langle v_1 \otimes v_2, w_1 \otimes w_2 \rangle \equiv \langle v_1, w_1 \rangle_1 \langle v_2, w_2 \rangle_2$$

for $v_1, w_1 \in H_1,\ v_2, w_2 \in H_2$.

Example 1.5 (Two-Qubit System)

By applying postulate 4, a two-qubit system can be represented in tensor product notation. Let us have qubits $\psi_0 = \alpha_0 |0\rangle + \beta_0 |1\rangle$ and $\psi_1 = \alpha_1 |0\rangle + \beta_1 |1\rangle$. Then, the tensor product notation of the two qubits can be represented as

$$|\psi_0\psi_1\rangle = \begin{pmatrix}\alpha_0\\ \beta_0\end{pmatrix} \otimes \begin{pmatrix}\alpha_1\\ \beta_1\end{pmatrix} = \begin{pmatrix}\alpha_0\alpha_1\\ \alpha_0\beta_1\\ \beta_0\alpha_1\\ \beta_0\beta_1\end{pmatrix},$$

with the important property that

$$|\alpha_0\alpha_1|^2 + |\alpha_0\beta_1|^2 + |\beta_0\alpha_1|^2 + |\beta_0\beta_1|^2 = 1,$$

where

$$\begin{aligned}\mathbb{P}(\text{collapse to } |00\rangle) &= |\alpha_0\alpha_1|^2,\\ \mathbb{P}(\text{collapse to } |01\rangle) &= |\alpha_0\beta_1|^2,\\ \mathbb{P}(\text{collapse to } |10\rangle) &= |\beta_0\alpha_1|^2,\\ \mathbb{P}(\text{collapse to } |11\rangle) &= |\beta_0\beta_1|^2.\end{aligned}$$

But since this tensor product space has the basis

$$|00\rangle = \begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix},\ |01\rangle = \begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix},\ |10\rangle = \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix},\ |11\rangle = \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix},$$

we can represent the two-qubit system more concisely as

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle,$$

with the measurement result $x(= 00, 01, 10, 11)$ occurring with probability $|\alpha_x|^2$.

We can now talk about entanglement.

Definition 1.8 (Entangled State)

A state $|\psi\rangle$ of a composite system that cannot be written as the tensor product of the states of its component systems is said to be in an **entangled state**.

Definition 1.9 (Bell States)

The four maximally entangled states of two qubits are known as the **Bell states**. They are

$$\left|\Phi^+\right\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \tag{18}$$

$$\left|\Psi^+\right\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}, \tag{19}$$

$$\left|\Phi^-\right\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}, \tag{20}$$

$$\left|\Psi^-\right\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}, \tag{21}$$

where the symbol represents the partity of the state (Φ for even and Ψ for odd) and the exponent represents the relative phase of the state ($+$ for positive and $-$ for negative).

Note that the Bell states cannot be written down as the following product

$$\begin{aligned}\frac{|00\rangle + |11\rangle}{\sqrt{2}} &\neq \big(\alpha|0\rangle + \beta|1\rangle\big) \otimes \big(\alpha'|0\rangle + \beta'|1\rangle\big) \\ &= \alpha\alpha'|00\rangle + \alpha\beta'|01\rangle + \beta\alpha'|10\rangle + \beta\beta'|11\rangle\end{aligned}$$

since this must mean that one of α or β' must be 0. If $\alpha = 0$, then the $|00\rangle$ amplitude must be 0 and if $\beta' = 0$, then the $|11\rangle$ amplitude must be 0, leading to a contradiction.

With further observation we can see that the majority of a composite Hilbert space consists of entangled states. That is, when we keep on adding qubits to the system, the total dimension grows as 2^n. However, the number of states that can be written as a tensor product of the individual qubits is only $2n$. Therefore, the majority of the states in the composite Hilbert space are entangled. This is similar to how most probability distributions have correlation, i.e. are not factorable into a product of independent distributions.

1.4.1 Bell's Inequality

Neilsen and Chuwang pg 118 and before.

1.5 Density Operators

So far, we've been doing things given that we *knew for sure what the state* $|\psi\rangle$ was. But what if we didn't know? We already had some uncertainty with measuring $|\psi\rangle$ with some observable, but now we have an additional layer of uncertainty on the state itself. This is where we can use density operators to represent *mixed states*.

Definition 1.10 (Density Operator)

Let i denote any set of index, and say that we have a quantum system in a state $|\psi_i\rangle$ with probability p_i. Then the **density operator** packages this ensemble of states into a single operator

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

If $|\{i\}| = 1$, then this is a **pure state**, and else it is a **mixed state**.

Theorem 1.4 (Properties of Density Operators)

The density operator has the following properties.

1. It has unit trace: $\text{Tr}(\rho) = 1$.
2. It is positive semidefinite: $\rho \geq 0$.
3. It is self-adjoint: $\rho^\dagger = \rho$.

Proof.

Listed.

1. By the cyclic trace property (just contraction of indices), we have
$$\text{Tr}(\rho) = \sum_i p_i \text{Tr}(|\psi_i\rangle \langle\psi_i|) = \sum_i p_i \langle\psi_i|\psi_i\rangle = \sum_i p_i = 1. \tag{22}$$
2. It is positive semidefinite since the eigendecomposition above shows that the eigenvalues are probabilities $p_i \geq 0$.
3. Again by the eigendecomposition, it is self-adjoint.

Theorem 1.5 (Criterion to Decide if a State is Mixed or Pure)

ρ is pure if

1. $\text{rank}(\rho) = 1$
2. ρ has 1 eigenvalue equal to 1, and rest equal to 0, which implies that $\rho^2 = \rho$ (is a projection operator).
3. $\text{Tr}(\rho^2) = 1$ (since cyclic trace property).

ρ is mixed if

1. $\text{rank}(\rho) > 1$
2. $\text{Tr}(\rho) = \sum_i p_i^2 < 1$ (since $\rho^2 \neq \rho$).

We've basically restated the state space postulate in terms of density operators. Now we can talk about their evolution and measurements. Recall that given a pure state $|\psi\rangle$, evolution is given by a unitary operator U acting on the state space. There is a natural extension for density operators.

Lemma 1.1 (Evolution of Density Operators)

Density operators evolve according to the rule
$$\rho \mapsto U\rho U^\dagger = \sum_i \rho_i U |\psi_i\rangle \langle\psi_i| U^\dagger \tag{23}$$
which makes sense since every $|\psi\rangle$ in the ensemble should also evolve unitarily according to U.

When we conduct a projective measurement, we have a random variable representing what we observe M and its associated Hamiltonian operator $\hat{M}$, the observable. We take the eigendecomposition of $\hat{M}$ into

its projection matrices $\hat{P}_m$ that project onto the eigenvectors. The probability of observing outcome m is

$$\mathbb{P}(M = m) = \langle\psi| \hat{P}_m |\psi\rangle \tag{24}$$

Lemma 1.2 (Measurement of Density Operators)

The above is really just like taking a trace of the 1×1 scalar generated, and so given a mixed state $\{|\psi_i\rangle\}$, the measurement on the ith state of the ensemble is

$$\mathbb{P}(M = m \mid i) = \text{Tr}(\langle\psi| \hat{P}_m |\psi\rangle) = \text{Tr}(\hat{P}_m |\psi_i\rangle \langle\psi_i|) \tag{25}$$

Using conditional probability and linearity, we can see that if we measure the mixed state ρ, the probability distribution of M will be

$$\mathbb{P}(M = m) = \sum_i \mathbb{P}(M = m \mid i)\mathbb{P}(i) = \sum_i \text{Tr}(\hat{P}_m |\psi_i\rangle \langle\psi_i|)\rho_i = \text{Tr}(\hat{P}_m \rho) \tag{26}$$

After the measurement, we want to take each state in the ensemble, project it into $\hat{P}_m$, and then renormalize it.

$$\rho \mapsto \rho' = \sum_i \mathbb{P}(i \mid M = m) \frac{\hat{P}_m |\psi_i\rangle \langle\psi_i| \hat{P}_m^\dagger}{\mathbb{P}(M = m \mid i)} \sum_i p(i) \frac{\hat{P}_m |\psi_i\rangle \langle\psi_i| \hat{P}_m}{\mathbb{P}(M = m)} = \frac{\hat{P}_m \rho \hat{P}_m}{\text{Tr}(\hat{P}_m \rho)} \tag{27}$$

Note that the trace of ρ' is still 1.

Lemma 1.3 (Expectation of Measurements on Density Operators)

Finally, the expectation value is simply

$$\mathbb{E}(M) = \sum_m m\mathbb{P}(M = m) = \sum_m m\text{Tr}(\hat{P}_m \rho) = \text{Tr}(\hat{M}\rho) \tag{28}$$

Let us go through some examples of density operators.

Example 1.6 (Simple Mixed State)

Consider the qubits and their density operators:

$$|\psi\rangle = |0\rangle \implies \rho = |0\rangle \langle 0| \tag{29}$$

$$|\psi\rangle = |1\rangle \implies \rho = |1\rangle \langle 1| \tag{30}$$

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \implies \rho = \frac{1}{2}\big(|0\rangle \langle 0| + |0\rangle \langle 1| + |1\rangle \langle 0| + |1\rangle \langle 1| \big) \tag{31}$$

These are all pure states,and by removing the middle two terms in the last operator, we can get a mixed state.

$$\rho = \frac{1}{2}\big(|0\rangle \langle 0| + |1\rangle \langle 1| \big) \tag{32}$$

which is a mixture $p_0 = p_1 = \frac{1}{2}$ of the pure states $|0\rangle$ and $|1\rangle$.

Exercise 1.1 ()

Show that both of the pure state mixtures ρ_1 and ρ_2 in the previous exercise can be created from the 2 qubit maximally entangled Bell state $|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$ by projectively measuring one qubit and

recombining the resulting single-qubit states into a single ensemble.

$$\rho_1 = \rho_2 = \sum_k (\langle e_k| \otimes I_2)\rho_\Phi + (|e_k\rangle \otimes I_2) \tag{33}$$

for $|e_k\rangle$ spanning two different measurement bases.

Exercise 1.2 ()

Imagine you are given two ensembles of single-qubit quantum states, one pure $|\psi\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$ and one mixed $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$.

1. Describe an experimental protocol to distinguish between the two ensembles.
2. What is the minimum possible (best-case) number of measurements you would need to conclusively distinguish between the two ensembles?

1.5.1 Ensemble Equivalence

You can see that the density matrix is not necessarily unique from the example above. Therefore, we should not assume some unique ensemble of states behind this density operator, but rather an equivalence class of states. A natural question to ask is what class of ensembles give rise to a single density matrix?

Definition 1.11 (Ensemble Equivalence)

Two ensembles $\{p_i, |\psi_i\rangle\}$ and $\{q_i, |\phi_i\rangle\}$ are said to be **ensemble equivalent** if they give rise to the same density operator.

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| = \sum_i q_i |\phi_i\rangle\langle\phi_i| \tag{34}$$

Theorem 1.6 (Ensemble Equivalence)

Let us have 2 ensembles $\{|\psi_i\rangle\}$ and $\{|\phi_i\rangle\}$ of size M and N, respectively, with $M \leq N$. Then, the two ensembles are ensemble equivalent if the following holds:

1. We first pad the smaller ensemble with 0 vectors to make them the same size.
2. There exists a unitary operator U such that $|\phi_i\rangle = U|\psi_i\rangle$

Consequently, two ensembles $\{p_i, |\psi_i\rangle\} and \{q_i, |\phi_i\rangle\}$ have the same density matrix if and only if they are ensemble equivalent in the above sense and U satisfies

$$\sqrt{p_i}|\psi_i\rangle = \sum_j U_i j \sqrt{q_j}|\phi_j\rangle \tag{35}$$

Proof.

TBD, Neilsen and Chuwang pg 104.

1.5.2 Pauli Decomposition of Density Operators

Recall the Pauli decomposition, which we can use to represent density operators.

Definition 1.12 (Pauli Decomposition)

The **Pauli Decomposition** is a generic decomposition for Hermitian 2×2 matrices. Since the Pauli matrices span the subspace of traceless Hermitian matrices, we can add in an identity to make the trace 1.

$$\rho(\mathbf{r}) = \frac{\mathbf{I} + \mathbf{r} \cdot \boldsymbol{\sigma}}{2} \tag{36}$$

where the $\frac{1}{2}$ constant normalizes the trace. It is set to be the dimension of the underlying Hilbert space (i.e. 2).

For pure states, we must have $\rho^2 = \rho$, and so

$$\rho^2 = \frac{1}{4}\left(\mathbf{I} + 2\mathbf{r} \cdot \boldsymbol{\sigma} + (\mathbf{r} \cdot \boldsymbol{\sigma})^2\right) \tag{37}$$

which is true if $||\mathbf{r}|| = 1$. This means that $\mathbf{r}$ must live on the unit sphere, i.e. the Bloch sphere.

Example 1.7 (Pauli Decomposition of $|0\rangle$)

If we have the pure state $|\psi\rangle = |0\rangle$, then

$$\rho = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \implies \mathbf{r} = \left(\frac{\text{Tr}(X\rho)}{2}, \frac{\text{Tr}(Y\rho)}{2}, \frac{\text{Tr}(Z\rho)}{2} \right) = (0, 0, 1) \tag{38}$$

If we have $|\psi\rangle = |+\rangle$, then

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \implies \mathbf{r} = (1, 0, 0) \tag{39}$$

For mixed states, we must have $\rho^2 < \rho$ (as in the $\rho - \rho^2$ is positive semidefinite), so some algebra reveals that

$$||\mathbf{r}|| < 1 \tag{40}$$

and so $\mathbf{r}$ must live inside the Bloch sphere.

Example 1.8 (Pauli Decomposition of Maximally Mixed State)

Given the ensembles $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$, the density operator is

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{41}$$

which we can see that $\mathbf{r} = (0, 0, 0)$, and so it is the center of the Bloch sphere.

1.5.3 Reduced Density Operator

We can also use the density operator to describe subsystems of a composite quantum system. To do this, we define the partial trace.

Definition 1.13 (Partial Trace)

Given a composite system on Hilbert space $V \otimes W$ and a linear map $A \otimes B : V \otimes W \to V \otimes W$, the **partial trace** of $A \otimes B$ is

$$\text{Tr}_B(A \otimes B) = A \cdot \text{Tr}(B) \tag{42}$$

which is just a tensor contraction over B. That is, by letting

$$B = \sum_{i,j} \beta_{ij} e_i \otimes e_j \implies \text{Tr}(B) = \sum_{i,j} \beta_{ij} e_i(e_j) = \sum_i \beta_{ii} \tag{43}$$

Definition 1.14 (Reduced Density Operator)

The **reduced density operator** of a composite system is the partial trace of the density operator of the composite system.

2 Quantum Circuits

By solving the time-independent Schrodinger's equation, we have found out that closed quantum systems evolve according to a unitary operator. A question to ask is what kind of unitary operators are natural to consider? In the case of single qubits, it turns out that *any* unitary operator at all can be realized in physical systems. Therefore, we can modify these qubits with general unitary operators, which are in some way a generalization of classical logic gates. Quantum logic gates can be interpreted as matrices that modify the state of qubits.

This is similar to those of classical gates, but one key difference between quantum logic gates and classical ones is that quantum gates are always invertible (from the unitary condition), while some classical gates like NAND are irreversible (e.g., if the output of a NAND gate is 1, we don't know if the input is $00, 01$, or 10). We can divide them into classes depending on how many arguments they take, but as we will see, the general form of a quantum gate taking in n input qubits is some unitary matrix in $U(2^n)$. Let us first focus on gates of one qubit. We will start with the Pauli matrices and spend some time investigating their properties.

1. Quantum circuits are acyclic, meaning that there are no loops, unlike classical circuits.
2. Classical circuits have the bits either being $|0\rangle$ or $|1\rangle$. Quantum circuits have the qubits being in a superposition $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$.
3. Classical circuits allow wires to be joined together (e.g. FANIN) or split/copied (e.g. FANOUT). Quantum circuits cannot do this since it is not invertible. Every gate must be a $n \times n$ unitary operator.

It is common to use both classical bits and qubits in a quantum circuit. To distinguish between the two, we use two lines to denote the flow of classical bits and a single line to denote the flow of qubits. This is clarified in the following examples.

2.1 Classical Logic Gates

Before we begin with quantum circuits, it is good to draw analogies to classical circuits. Recall the following below, with notation changed to quantum mechanical notation.

Definition 2.1 (Classical NOT)

The not gate is

Definition 2.2 (Classical AND)

Definition 2.3 (Classical XOR)

In the general case, the XOR gate maps

$$\text{XOR}: \ (A, B) \mapsto (A, A \oplus B) \tag{44}$$

where A and B are classical bits. This can be used as a FANOUT gate, and so we can use it to copy a bit.

2.2 Quantum Logic Gates

Definition 2.4 (Pauli X)

The Pauli X gate is the quantum equivalent of the classical NOT gate.

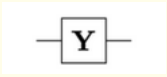
1. It transforms as such:
$$\alpha |0\rangle + \beta |1\rangle \mapsto \beta |0\rangle + \alpha |1\rangle \tag{45}$$
2. with the matrix form:
$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{46}$$
3. and the circuit form:

4. It also flips the state of the qubit in the Bloch sphere about the X-axis.

Definition 2.5 (Pauli Y)

The Pauli Y gate

1. It transforms as such:
$$\alpha |0\rangle + \beta |1\rangle \mapsto i\beta |0\rangle - i\alpha |1\rangle \tag{47}$$
2. with the matrix form:
$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{48}$$
3. and the circuit form:



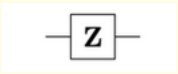
4. It also flips the state of the qubit in the Bloch sphere about the Y-axis.

Definition 2.6 (Pauli Z)

The Pauli Z gate simply flips the sign of the coefficient of $|1\rangle$.

1. It transforms as such:
$$\alpha |0\rangle + \beta |1\rangle \mapsto \alpha |0\rangle - \beta |1\rangle \tag{49}$$

2. with the matrix form:
$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{50}$$
3. and the circuit form:



4. It also flips the state of the qubit in the Bloch sphere about the Z-axis.

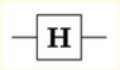
Since the Pauli matrices form a basis for the space of 2×2 Hermitian matrices, any unitary operator acting on a single qubit can be written as a linear combination of the Pauli matrices. By setting $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ to be the vector of Pauli matrices, we can write any unitary operator as

$$\mathbf{v} \cdot \boldsymbol{\sigma} = v_1\sigma_1 + v_2\sigma_2 + v_3\sigma_3 \tag{51}$$

Definition 2.7 (Hadamard)

The **Hadamard gate** H takes in $|0\rangle$ or $|1\rangle$ and puts it into exactly equal superposition. That is, $H|0\rangle = |+\rangle$ and $H|1\rangle = |-\rangle$. Furthermore, it can take Bell states and put them into either $|0\rangle$ or $|1\rangle$.

1. It transforms as such:
$$\alpha\,|0\rangle + \beta\,|1\rangle \mapsto \frac{\alpha+\beta}{\sqrt{2}}\,|0\rangle + \frac{\alpha-\beta}{\sqrt{2}}\,|1\rangle \tag{52}$$
2. with the matrix form
$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{53}$$
3. and the circuit form



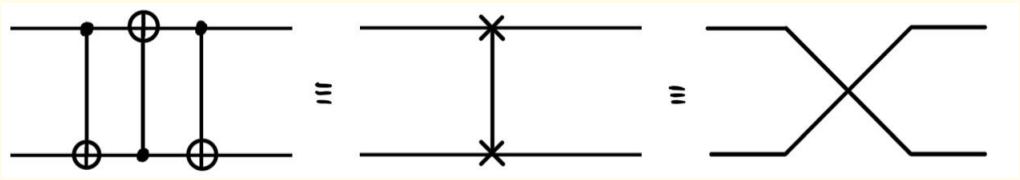
4. This can be thought of as a rotation around the Bloch vector $(1, 0, 1)$.

The next set consists of multiple qubit gates. Recall that any function of bits can be computed from the composition of NAND gates alone, which is known as a **universal gate**. The multi-qubit universal quantum gate is actually the control-not gate.

Definition 2.8 (Swap)

The **swap** gate simply swaps the states of the two qubits.

1. It transforms as such:
$$a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle \mapsto a|00\rangle + c|01\rangle + b|10\rangle + d|11\rangle \tag{54}$$
2. with the matrix form
$$U_{SWAP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{55}$$
3. and the circuit form



Definition 2.9 (Control NOT)

The **controlled-NOT** is a 2 qubit gate that has a control qubit. If the control qubit is set to $|0\rangle$, then the target qubit is left alone. If the control qubit is set to $|1\rangle$, then the target qubit is flipped. This seems a bit confusing since a qubit can be neither $|0\rangle$ nor $|1\rangle$, but this rule applies to each component of the tensor product.

1. It transforms as such:

$$a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle \mapsto a|00\rangle + b|01\rangle + d|10\rangle + c|11\rangle \tag{56}$$

2. with the matrix form:

$$U_{CNOT} = \begin{pmatrix} I & 0 \\ 0 & X \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{57}$$

3. and the circuit form

Notice how unlike the swap gate, the output 4-vector of the CNOT gate is not always guaranteed to decompose into a tensor product $|\psi_1\rangle \otimes |\psi_2\rangle$ of vectors (more on Bell states later). Its circuit is represented in the given image.

At this point, we can now consider an extremely useful application: generating entangled states. Essentially, given two qubits $|00\rangle$, we can use a Hadamard gate to put the first qubit into a superposition, and then use a CNOT gate to introduce some correlation, ultimately entangling the two qubits.

Lemma 2.1 (Entangling Qubits)

Given a two qubit system, we can use a Hadamard gate and a CNOT gate to entangle the two qubits as such:

1. Apply a Hadamard gate to the first qubit.
2. Apply a CNOT gate with the first qubit as the control and the second qubit as the target.

By doing this produre to the four computational basis states $|00\rangle, |01\rangle, |10\rangle, |11\rangle$, can get get all four Bell states.

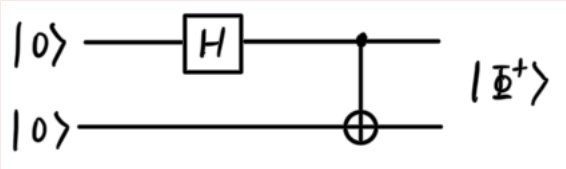


Figure 1: Entangling procedure done for the $|00\rangle$ basis state to produce $|\Phi^+\rangle$.

Furthermore, we can just start off with the $|00\rangle$ state and apply some unitary gate U on the first qubit to get all 4 Bell states.

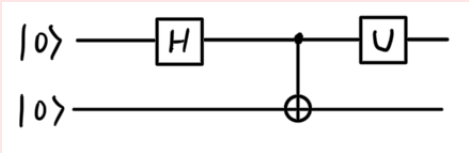


Figure 2: Entangling procedure done for the $|00\rangle$ basis state to produce any of the 4 Bell states. Setting $U = I, X, Z, ZX$ gives the bell states $|\Phi^+\rangle, |\Psi^+\rangle, |\Phi^-\rangle, |\Psi^-\rangle$.

Definition 2.10 (Control U)

The **controlled-U** gate is a generalization of controlled-NOT. Let us have a control bit and n target bits. If the control bit is set to $|0\rangle$, then the target qubits are left alone. If the control qubit is set to $|1\rangle$, then the states/spins of the n target qubits are changed by some unitary matrix $U \in \mathrm{U}(2^n)$.

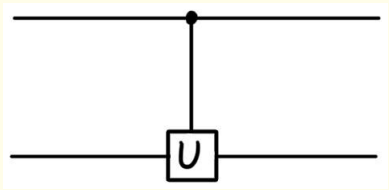
1. It transforms as such:

$$a\,|0\rangle \otimes |\psi\rangle + b\,|1\rangle \otimes |\psi\rangle \mapsto a\,|0\rangle \otimes |\psi\rangle + b\,|1\rangle \otimes U\,|\psi\rangle \tag{58}$$

2. with the matrix form:

$$U_{CU} = \begin{pmatrix} I & 0 \\ 0 & U \end{pmatrix} \tag{59}$$

3. and the circuit form:



Definition 2.11 (Toffoli)

The **Toffoli** gate is similar to a CNOT but with two control qubits and 1 target qubit. If the control qubits are set to $|11\rangle$, then the target qubit is flipped.

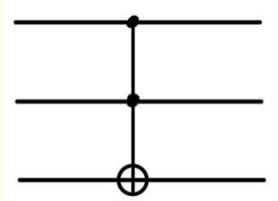
1. It transforms as such:

$$\begin{aligned} &a|000\rangle + b|001\rangle + c|010\rangle + d|011\rangle + e|100\rangle + f|101\rangle + g|110\rangle + h|111\rangle \\ &\quad \mapsto a|000\rangle + b|001\rangle + c|010\rangle + d|011\rangle + e|100\rangle + f|101\rangle + h|110\rangle + g|111\rangle. \end{aligned}$$

2. with the matrix form:

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{60}$$

3. and the circuit form:



At this point, we can see that we can have an arbitrary number of qubits which act as a control for some unitary operator. This gives us great flexibility in constructing quantum circuits.

Definition 2.12 (Measurement)

Finally, we introduce the operation of "measurement," which we represent by a meter symbol.

This operation converts a single qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ into a probabilistic classical bit M, which is 0 with probability $|\alpha|^2$ or 1 with probability $|\beta|^2$.

2.3 Physical Implementation

3 Quantum Computation and Properties

3.1 Quantum Tomography

In quantum computing, uncertainty is a huge obstacle to overcome, and it is not always the case that we know which state we are working with in a qubit. In a classical computer, we can always measure the state of a bit[1] However, this is much harder in quantum computing. It is not so simple to just "see" which state a qubit is in, but it is possible. We typically do this through measurements, but as we know, measurements collapse the state of the qubit. Therefore, we want to measure some qubit (or an ensemble of identical qubits) and then reconstruct the state of the qubit according to the information that we have gained from our qubits.

Definition 3.1 (Quantum Tomography)

Quantum tomography refers to the process of determining the state of some system by measuring a system of identical systems. That is, while measurements tell you the state of the system after

[1]You can simply measure the electrical pulse going through a wire. Presence indicates that it is turned on, a 1, and the absence indicates that it is off, a 0.

the measurement, tomography gives you insight about what the state was *before* the measurement. Then, with this system we can reconstruct the state of the qubit.[a]

[a]This can be done by constructing the correct gate G and just putting a bunch of $|0\rangle$ states through it.

In the end, it is just a question of *what is the maximal amount of information I can extract from this system to reconstruct it?*. Sometimes, this is not possible to know for certain, but we can say with high probability. This is very similar to the problem of statistics, and in a way it is. We essentially have some distribution, and every measurement samples from that distribution. Given these samples, we want to reconstruct the distribution, and techniques like maximum likelihood or Bayesian updating are even used. But let's not get ahead of ourselves and start with an example. There are many assumptions that get thrown around, listed below, and we should know the basic ones.

1. Whether we are trying to determine the state of a single qubit or a system of multiple qubits.
2. Whether we know a set of state $\{|\psi_i\rangle\}$ that the true state is in.
3. Whether these states are orthonormal.
4. How many qubits are prepared for us to work with?[2]

Example 3.1 (Distinguishing Two Orthonormal States)

You are given a single qubit $|\psi\rangle$ which can be in one of two orthonormal state $|\psi_1\rangle$ and $|\psi_2\rangle$. It turns out that you can perfectly distinguish these states.

1. Construct a measurement operator

$$\hat{M} = 1 \cdot \hat{P}_1 + 2 \cdot \hat{P}_2 = 1 \cdot |\psi_1\rangle \langle\psi_1| + 2 \cdot |\psi_2\rangle \langle\psi_2| \tag{61}$$

2. When you measure the qubit $|\psi\rangle$, it must be an eigenvector of $\hat{M}$. Let p be the random variable of the prepared state. Then

$$\mathbb{P}(M = 1 \mid p = |\psi_1\rangle) = \langle\psi_1| \hat{P}_1 |\psi_1\rangle = 1 \tag{62}$$
$$\mathbb{P}(M = 1 \mid p = |\psi_2\rangle) = \langle\psi_2| \hat{P}_1 |\psi_2\rangle = 0 \tag{63}$$
$$\mathbb{P}(M = 2 \mid p = |\psi_1\rangle) = \langle\psi_1| \hat{P}_2 |\psi_1\rangle = 0 \tag{64}$$
$$\mathbb{P}(M = 2 \mid p = |\psi_2\rangle) = \langle\psi_2| \hat{P}_2 |\psi_2\rangle = 1 \tag{65}$$

3. Therefore, if you observe $M = 1$, then it must have been the case that the state was $|\psi_1\rangle$, and if you observe $M = 2$, then it must have been the case that the state was $|\psi_2\rangle$. To formalize this, it is just simply Bayes rule.

$$\mathbb{P}(p = |\psi_1\rangle \mid M = 1) = \frac{\mathbb{P}(M = 1 \mid p = |\psi_1\rangle)\mathbb{P}(p = |\psi_1\rangle)}{\mathbb{P}(M = 1)} = 1 \tag{66}$$
$$\mathbb{P}(p = |\psi_2\rangle \mid M = 2) = \frac{\mathbb{P}(M = 2 \mid p = |\psi_2\rangle)\mathbb{P}(p = |\psi_2\rangle)}{\mathbb{P}(M = 2)} = 1 \tag{67}$$

We can see that depending on what the measurement outcome realizes to, there is a probability of us being certain of the state. Let's define this metric.

[2]Note that this must be finite, since if it was infinite, then we can measure with infinite precision and our state is trivially constructed.

Definition 3.2 (Probability of Certainty)

The **probability of certainty** is the probability of some measurement outcome M being realized that gives us certainty of the state.

Therefore, the probability of certainty of the example above was 1. Okay, that was easy, but what if the states were not orthonormal? It turns out that it is impossible to perfectly distinguish between two non-orthonormal states, even if we have an unbounded but finite number of identically prepared states to work with.

Lemma 3.1 (Distinguishing Non-Orthonormal States)

Given a finite number of n qubits identically prepared at state $|\psi\rangle$ which can take states within the set $\{|\psi_i\rangle\}$, it is not possible to perfectly distinguish them if the states are not orthonormal.

Proof.

Suppose it is possible, corresponding to the measurement M with its Hermitian operator $\hat{M}$. Then, it should be the case that if we start with state $|\psi_1\rangle$, then $\mathbb{P}(M=1)=1$, meaning that $|\psi_1\rangle$ must be an eigenvector of M. This holds true for $|\psi_2\rangle$ and so it must also be an eigenvector. However, this violates the contrapositive of the Spectral theorem which states that *if A is Hermitian on V, then there exists an orthonormal basis of V consisting of eigenvectors of A.*

We can still distinguish them with high probability the more prepared states we have, but before blindly doing this, we should try to maximally optimize each measurement. Note that there are three stages to this process:

1. Prepare the state $|\psi\rangle$.
2. Measure the state $|\psi\rangle$ with some measurement operator $\hat{M}$.
3. Guess the state $|\psi\rangle$ based on the measurement outcome.

The user can control the measurement operator and separately determine a strategy for guessing the state. We have implicitly made a strategy above using basic probability, but you can make completely bogus strategies even if you have the optimal measurement operator. With this in mind, we define more fine metrics.

Definition 3.3 ((Maximal) Probability of Success)

Given some measurement operator $\hat{M}$ and some strategy $\mathcal{S}$, the **probability of success** is the probability that the strategy $\mathcal{S}$ correctly guesses the state $|\psi\rangle$ given the measurement outcome M. The **maximal probability of success** is the maximum probability of success over all strategies. Note that for such thing to exist we must place a prior distribution over the states (i.e. know the probability of each state being prepared).

Example 3.2 (Distinguishing Two Non-Orthonormal States)

Suppose we have a single $|\psi\rangle$ qubit which can be in one of two states $|\psi_1\rangle = |0\rangle$ and $|\psi_2\rangle = |+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$, prepared with equal probability. We can naively construct the measurement operator

$$\hat{M} = 1 \cdot \hat{P}_1 + 2 \cdot \hat{P}_2 = 1 \cdot |0\rangle\langle 0| + 2 \cdot |1\rangle\langle 1| \tag{68}$$

By focusing on $|0\rangle$ and setting the other one to be $|0^\perp\rangle$. We know this won't be perfect, but hopefully it will give us a decent chance of predicting correctly. We want to calculate both the probability of

certainty and the maximal probability of success. First, let's compute some probabilities.

$$\mathbb{P}(M=1 \mid p=|0\rangle) = \langle 0| \hat{P}_1 |0\rangle = 1 \tag{69}$$

$$\mathbb{P}(M=1 \mid p=|+\rangle) = \langle +| \hat{P}_1 |+\rangle = \frac{1}{2} \tag{70}$$

$$\mathbb{P}(M=2 \mid p=|0\rangle) = \langle 0| \hat{P}_2 |0\rangle = 0 \tag{71}$$

$$\mathbb{P}(M=2 \mid p=|+\rangle) = \langle +| \hat{P}_2 |+\rangle = \frac{1}{2} \tag{72}$$

Therefore, with Bayes rule, we can compute

$$\mathbb{P}(p=|0\rangle \mid M=1) = \frac{2}{3} \tag{73}$$

$$\mathbb{P}(p=|+\rangle \mid M=1) = \frac{1}{3} \tag{74}$$

$$\mathbb{P}(p=|0\rangle \mid M=2) = 0 \tag{75}$$

$$\mathbb{P}(p=|+\rangle \mid M=2) = 1 \tag{76}$$

and the probability of certainty is

$$\mathbb{P}(M=2) = \mathbb{P}(M=2 \mid p=|0\rangle)\mathbb{P}(p=|0\rangle) + \mathbb{P}(M=2 \mid p=|1\rangle)\mathbb{P}(p=|1\rangle) \tag{77}$$

$$= 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \tag{78}$$

With this measurement operator, our optimal strategy would be to just simply choose the state with the maximal probability in the posterior. If $M=1$, we should choose $|0\rangle$ and if $M=2$, we should choose $|+\rangle$. The maximal probability of success is

$$\mathbb{P}(M=1 \mid p=|0\rangle)\mathbb{P}(p=|0\rangle) + \mathbb{P}(M=2 \mid p=|+\rangle)\mathbb{P}(p=|+\rangle) = \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{3}{4} \tag{79}$$

Determining the best strategy is pretty easy, but what about the best operator? Well this is simply just an optimization problem.

Example 3.3 (Best Operator for Distinguishing Two Non-Orthonormal States)

Given the same situation above, let's be a bit smarter about constructing $\hat{M}$. It must be defined over some orthonormal basis, so we can parameterize it from the Bloch sphere:

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle \tag{80}$$

$$|\psi^\perp\rangle = e^{-i\phi}\sin\frac{\theta}{2}|0\rangle - \cos\frac{\theta}{2}|1\rangle \tag{81}$$

and so our measurement is

$$\hat{M} = 1 \cdot \hat{P}_1 + 2 \cdot \hat{P}_2 = 1 \cdot |\psi\rangle\langle\psi| + 2 \cdot |\psi^\perp\rangle\langle\psi^\perp| \tag{82}$$

Therefore, we can compute with some trigonometric identities

$$\mathbb{P}(M=1 \mid p=|0\rangle) = \langle 0|\hat{P}_1|0\rangle = \cos^2\frac{\theta}{2} \tag{83}$$

$$\mathbb{P}(M=1 \mid p=|+\rangle) = \langle +|\hat{P}_1|+\rangle = \frac{1}{2}+\frac{1}{2}\sin\theta\cos\phi \tag{84}$$

$$\mathbb{P}(M=2 \mid p=|0\rangle) = \langle 0|\hat{P}_2|0\rangle = \sin^2\frac{\theta}{2} \tag{85}$$

$$\mathbb{P}(M=2 \mid p=|+\rangle) = \langle +|\hat{P}_2|+\rangle = \frac{1}{2}-\frac{1}{2}\sin\theta\cos\phi \tag{86}$$

Notice that computing the inverse conditionals with Bayes rule isn't necessary. These probabilities are only required for us to compute the probability of certainty, but we know that as long as the optimal basis does not coincide with the possible states, this will be 0. Let's directly go into calculating the maximal probability of success, which is

$$\mathbb{P}(\text{success}) = \mathbb{P}(M=1 \mid p=|0\rangle)\mathbb{P}(p=|0\rangle) + \mathbb{P}(M=2 \mid p=|+\rangle)\mathbb{P}(p=|+\rangle) \tag{87}$$

$$= \cos^2\frac{\theta}{2}\cdot\frac{1}{2} + \left(\frac{1}{2}-\frac{1}{2}\sin\theta\cos\phi\right)\cdot\frac{1}{2} \tag{88}$$

We want to maximize this, and first we can remove all the constant terms. Second, we can see that maximizing w.r.t. ϕ is independent of θ, and so we can set $\phi = 0$. Then, we can maximize w.r.t. θ to get the maximal probability of success. We just set the derivative to 0 and solve.

$$0 = \frac{d}{d\theta}\left\{\cos^2\frac{\theta}{2} - \frac{1}{2}\sin\theta\right\} \implies \theta = \frac{\pi}{4} \tag{89}$$

Therefore, our optimal measurement basis is

$$|\psi\rangle = \cos\frac{\pi}{8}|0\rangle + \sin\frac{\pi}{8}|1\rangle \tag{90}$$

and our maximal probability of success is

$$\mathbb{P}(\text{success}) = \frac{1}{2}+\frac{1}{2\sqrt{2}} \tag{91}$$

Geometric interpretation.

Theorem 3.1 ()

The maximum probability of success to distinguish two states $|\psi_1\rangle, |\psi_2\rangle$ is

$$\mathbb{P}(\text{success}) = \frac{1+\sin(\arccos(|\langle\psi_1|\psi_2\rangle|))}{2} \tag{92}$$

Theorem 3.2 (Holevo Bound)

The Holevo bound is a generalization of the previous theorem to n states.

$$\mathbb{P}(\text{success}) \leq 1 - H\left(\frac{1+|\langle\psi_1|\psi_2\rangle|}{2}\right) \tag{93}$$

where H is the binary entropy function.

3.2 Parity Operators

Quantum tomography is the first tool we need to work with states, but we have said that often we need an ensemble of identical qubits to work with so we can make an optimal series of measurements. This is needed since these measurements are destructive, and we need to make sure that we can extract as much information as possible. But every measurement does not necessarily have to be destructive. We can use *partial projective measurements* to extract information while still keeping it in some superposition by projecting the system into some non-degenerate eigenspace of our measurement operator $\hat{M}$.

Example 3.4 (Computing Parity)

Say that we have a n-qubit state $|\psi\rangle$ and we want to measure the parity of the state. There are two ways that we can do this:

1. We can perform a complete Von Neumann measurement with respect to the computational basis and subsequently compute the parity of the resulting string.
2. We can perform a projective measurement of the parity with the operator $Z^{\otimes n}$

They both measure the parity of the realized system, but the difference is what the system collapses to. Let's have a state $|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$. Then, I can take an ancilla qubit in the state $|0\rangle$ and conduct the parity measurement

$$U_p = (|00\rangle\langle 00| + |11\rangle\langle 11|) \otimes I + (|01\rangle\langle 01| + |10\rangle\langle 10|) \otimes X \tag{94}$$

which brings the state to

$$|\psi\rangle \otimes |0\rangle \xrightarrow{U_p} (\alpha_{00}|00\rangle + \alpha_{11}|11\rangle) \otimes |0\rangle + (\alpha_{01}|01\rangle + \alpha_{10}|10\rangle) \otimes |1\rangle \tag{95}$$

By conducting a partial projective measurement on the third qubit, we are essentially projecting the state to the eigenspace corresponding to the parity of the state, which is spanned by either $\{|00\rangle + |11\rangle\}$ and $\{|01\rangle + |10\rangle\}$. Therefore, the second measurement may be preferred since we can directly observe parity without destroying the superposition of the system.

We can get a good geometric interpretation of this. Initially, there is some quantum system $|\psi\rangle$ in some Hilbert space $\mathcal{H}$. Every projective measurement, complete or partial, projects $|\psi\rangle$ into some lower dimensional subspace, and by projecting it we can glean information about it. The price to pay is that after every projection, some information about the system is lost, and after all projections, we know that the system must be in some 1-dimensional subspace, where all information is lost.

3.3 No Cloning Theorem

Let us step back into the world of classical computing with classical bits. It is important to know that the CNOT and Toffoli gates take in *quantum bits*, not just classical ones.

1. The classical analogue of the CNOT gate is the XOR gate (or mathematically speaking, XOR is the restriction of CNOT to qubits in superposition of $|0\rangle$ or $|1\rangle$). One of the reasons that XOR is significant is because it can be used as a FANOUT gate, which takes in an arbitrary classical bit 0 or 1 and essentially copies it to output 00 or 11, where each copy can be used for separate purposes. By setting the first bit to be some arbitrary $A \in \{|0\rangle, |1\rangle\}$ that we want to copy and the second bit $B = 0$ as a "scratchpad" bit, we have

   $$\text{XOR}: \ (A, 0) \mapsto (A, A \oplus 0 = A) \tag{96}$$

   To explicitly calculate, we can do

   $$U_{\text{CNOT}}(|0\rangle \otimes |0\rangle) = |0\rangle \otimes |0\rangle \tag{97}$$
   $$U_{\text{CNOT}}(|1\rangle \otimes |0\rangle) = |1\rangle \otimes |1\rangle \tag{98}$$

2. The Toffoli gate can also be used as a FANOUT gate to copy classical bits since given classical bits A, B, C, the Toffoli gate maps

$$\text{Toffoli}: \ (A, B, C) \mapsto (A, B, C \oplus AB) \tag{99}$$

where A, B, C are classical bits. By setting the first bit $A = 1$, B as an arbitrary bit to clone, and $C = 0$, we have

$$\text{Toffoli}: \ (1, B, 0) \mapsto (1, B, 0 \oplus 1 \cdot B = B) \tag{100}$$

We have just demonstrated that it is very much possible to clone a classical bit 0 or 1, i.e. a quantum bit in a superposition of $|0\rangle$ or $|1\rangle$. However, it turns out that we cannot copy a qubit in some general state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. Let us naively apply this for a general qubit $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$. We get

1. For a CNOT gate, we get

$$\begin{aligned} U_{\text{CNOT}}(|\psi\rangle \otimes |0\rangle) &= U_{\text{CNOT}}\big((\alpha\,|0\rangle + \beta\,|1\rangle) \otimes |0\rangle\big) && (101)\\ &= U_{\text{CNOT}}(\alpha\,|00\rangle + \beta\,|10\rangle) && (102)\\ &= \alpha\,|00\rangle + \beta\,|11\rangle && (103)\\ &\neq \alpha^2\,|00\rangle + \alpha\beta\,|10\rangle + \alpha\beta\,|01\rangle + \beta^2\,|11\rangle && (104)\\ &= (\alpha\,|0\rangle + \beta\,|1\rangle) \otimes (\alpha\,|0\rangle + \beta\,|1\rangle) && (105)\\ &= |\psi\rangle \otimes |\psi\rangle && (106) \end{aligned}$$

2. For a Toffoli gate by inputting $|1\rangle \otimes |\psi\rangle \otimes |0\rangle$, we get

$$T\begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}\alpha\\\beta\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix} = T\begin{pmatrix}0\\0\\0\\0\\\alpha\\0\\\beta\\0\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\\\alpha\\0\\0\\\beta\end{pmatrix} = \alpha|100\rangle + \beta|111\rangle \tag{107}$$

which is not equal to $|1\rangle \otimes |\psi\rangle \otimes |\psi\rangle = \alpha^2|000\rangle + \alpha\beta|001\rangle + \alpha\beta|010\rangle + \beta^2|011\rangle$. No copying is done.

We may try to experiment with other schemes, but it will be fruitless. We will show that this is theoretically impossible in quantum computing. This is not the most general proof since we only consider 2-qubit operators, but it can extend beyond this.

Theorem 3.3 (No Cloning Theorem)

There exists no 2-qubit gate L such that for any qubit state $|\psi\rangle$, we have

$$L(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle. \tag{108}$$

Proof.

Assume that there exists such a gate L. Then, it should be that

$$\begin{aligned} L(|\psi\rangle \otimes |0\rangle) &= |\psi\rangle \otimes |\psi\rangle && (109)\\ &= (\alpha\,|0\rangle + \beta\,|1\rangle) \otimes (\alpha\,|0\rangle + \beta\,|1\rangle) && (110)\\ &= \alpha^2\,|00\rangle + \alpha\beta\,|01\rangle + \alpha\beta\,|10\rangle + \beta^2\,|11\rangle && (111) \end{aligned}$$

We can also see that

$$|\psi\rangle \otimes |0\rangle = \alpha\,|00\rangle + \beta\,|10\rangle \tag{112}$$

But by linearity of L, we have

$$L(\alpha|00\rangle + \beta|10\rangle) = \alpha L(|00\rangle) + \beta L(|10\rangle) \tag{113}$$
$$= \alpha|00\rangle + \beta|11\rangle \tag{114}$$
$$\tag{115}$$

which doesn't align with our previous equation. Therefore, no such L exists.

3.4 Propagating Entanglement

We've talked about entangling 2 qubits, which was relatively simple using the CNOT and Hadamard gates, but what if we wanted more?

Definition 3.4 (Greenberger-Horne-Zellinger (GHZ) State)

The **Greenberger-Horne-Zellinger (GHZ)** state is a type of entangled state of 3 qubits. It is defined as

$$|\text{GHZ}\rangle = \frac{|000\rangle + |111\rangle}{\sqrt{2}} \tag{116}$$

Theorem 3.4 (Generating the GHZ State from Bell States)

We can generate this from 2 qubit Bell states along with some assumptions. Given three parties A, B, C, we have the following:

1. A Bell state is generated, with one qubit given to A and the other qubit given to B.
2. A Bell state is generated, with one qubit given to B and the other qubit given to C.

A circuit diagram is first shown:

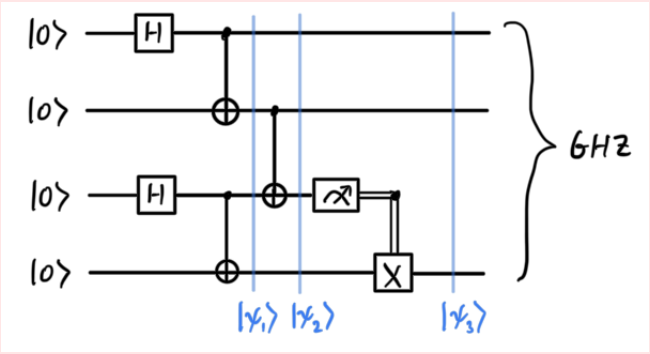


Figure 3: Circuit diagram for generating the GHZ state from Bell states.

By generating the GHZ state, we are essentially propagating the entanglement from A to C.

1. We start off with the state

$$|\psi_1\rangle = \left|\Phi^+\right\rangle \left|\Phi^+\right\rangle = \frac{|0000\rangle + |0011\rangle + |1100\rangle + |1111\rangle}{2} \tag{117}$$

2. Now we want to make a measurement on only 1 qubit to turn this into a 3 qubit system. The measurement operator that we can use is one that computes whether the 2 qubits that B has

match onto the third qubit. This can be done with the parity operator done on B's qubits, as you can see in the following.

$$|00\rangle \mapsto |0\mathbf{0}\rangle \tag{118}$$
$$|01\rangle \mapsto |1\mathbf{1}\rangle \tag{119}$$
$$|10\rangle \mapsto |0\mathbf{1}\rangle \tag{120}$$
$$|11\rangle \mapsto |1\mathbf{0}\rangle \tag{121}$$

This turns out to be the CNOT with the 2nd qubit being the control and the 3rd being the target.

$$|\psi_2\rangle = \frac{|0000\rangle + |0011\rangle + |1110\rangle + |1101\rangle}{2} \tag{122}$$

3. Now note that the first and fourth terms have the same third qubit of 0 and the second/third have the same third qubit of 1. If we measure the third qubit and it turns out to be 0, the entangled system will collapse to the first and fourth terms, which is the GHZ state. If it turns out to be 1, it will collapse to the second and third terms.

$$0 \text{ measured} \implies \psi_3^* = \frac{|000\rangle + |111\rangle}{\sqrt{2}} = |\text{GHZ}\rangle \tag{123}$$
$$1 \text{ measured} \implies \psi_3^* = \frac{|001\rangle + |110\rangle}{\sqrt{2}} \tag{124}$$

The second is not the GHZ state but can be turned into one by simply swapping the third qubit between 0 and 1. Since this is conditioned on the fact that the measured third (previous) qubit is 1, we can use a controlled gate with the X since we want them to negated (NOT). Therefore,

$$|\psi_3\rangle = \frac{|000\rangle + |111\rangle}{\sqrt{2}} = |\text{GHZ}\rangle \tag{125}$$

Question 3.1 (To Do)

Class 4, how do I create a 4-qubit GHZ state?

3.5 Quantum Teleportation

We have just proved that we cannot copy qubits, but what is possible is to transfer the state of one qubit to another. This is known as **quantum teleportation**, or more specifically **entanglement-assisted teleportation**. More specifically, suppose A has a qubit in quantum state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ and wants to send this qubit to B (by "sending" it to B, we mean that we want B to be in possession of a qubit in state $|\psi\rangle$ in some way). This is possible under very special circumstances, which we will describe in the theorem below.

Theorem 3.5 (Quantum Teleportation)

Given that two parties A wants to send B a qubit $|\psi\rangle$, it is possible to do so under the following assumptions:

1. An EPR pair (two qubits in some Bell state) has been generated with each qubit given to A and B. This can be done beforehand by A, B, or even a third party C.
2. A can communicate to B by sending classical information to B (i.e. finite strings of 0 and 1). The finiteness of this condition is most restrictive, since if A could send infinite strings, A can

just send the infinite binary representation of $|\psi\rangle$.

The entire circuit is represented by the diagram below:

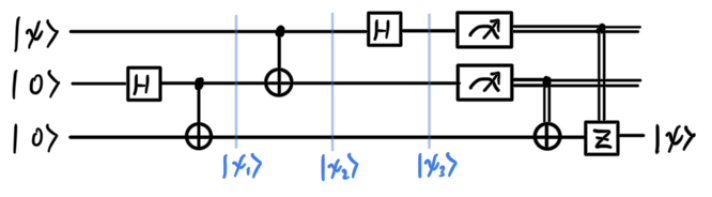


Figure 4: Quantum teleportation circuit, with A having the first and second qubits and B having the third qubit.

The following steps are taken:

1. Before the experiment, we take the $|00\rangle$ qubit and entangle it, bringing us to $|\psi_1\rangle$.

$$|\psi_1\rangle = |\psi\rangle \left|\Phi^+\right\rangle \tag{126}$$
$$= \frac{1}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle)(|00\rangle + |11\rangle) \tag{127}$$
$$= \frac{1}{\sqrt{2}}(\alpha|000\rangle + \alpha|011\rangle + \beta|100\rangle + \beta|111\rangle) \tag{128}$$

2. Then, we put take the first and second qubit through a control NOT gate. Since the control is the first qubit, we can just swap the second qubit from 0 to 1 or 1 to 0 if the first qubit is a 1 for every component.

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}(\alpha|000\rangle + \alpha|011\rangle + \beta|110\rangle + \beta|101\rangle) \tag{129}$$

3. Then, we put the first qubit through a Hadamard gate, which after some computation give us

$$\begin{aligned}
|\psi_3\rangle &= \frac{1}{2}\left(\frac{\alpha}{\sqrt{2}}(|0\rangle + |1\rangle)|00\rangle + \frac{\alpha}{\sqrt{2}}(|0\rangle + |1\rangle)|11\rangle + \frac{\beta}{\sqrt{2}}(|0\rangle - |1\rangle)|10\rangle + \frac{\beta}{\sqrt{2}}(|0\rangle - |1\rangle)|01\rangle\right) \\
&= \frac{1}{2}\left(\alpha(|000\rangle + |100\rangle + |011\rangle + |111\rangle +) + \beta(|010\rangle - |110\rangle + |001\rangle - |101\rangle)\right) \\
&= \frac{1}{2}\left(|00\rangle(\alpha|0\rangle + \beta|1\rangle) + |01\rangle(\alpha|1\rangle + |0\rangle) + |10\rangle(\alpha|0\rangle - \beta|1\rangle) + |11\rangle(\alpha|1\rangle - \beta|0\rangle)\right) \\
&= \frac{1}{2}\left(|00\rangle(I|\psi\rangle) + |01\rangle(X|\psi\rangle) + |10\rangle(Z|\psi\rangle) + |11\rangle(XZ|\psi\rangle)\right)
\end{aligned}$$

4. At this point $|\psi_3\rangle$ stores the state of $|\psi\rangle$, but there needs to be a bit of postprocessing, which can be done with the next gate that takes in the first two classical bits as the control. Depending on what classical bits we measure from the first and second bits, we can undo the operations we did to the third qubit.
   (a) If we measure 00, then we have $I|\psi\rangle$, and we don't need to do anything.
   (b) If we measure 01, then we have $X|\psi\rangle$, and we can undo this by applying X again.
   (c) If we measure 10, then we have $Z|\psi\rangle$, and we can undo this by applying Z again.
   (d) If we measure 11, then we have $XZ|\psi\rangle$, and we can undo this by applying X first and then Z.

Note that A must communicate to B the measurement outcome of the first two bits over a classic communication channel in order to complete the teleportation. Since this classic information is subject to the limits of speed of light, the teleportation of a qubit does not violate the upper limit. This example may also look like it has violated the No-Cloning theorem, since we have copied the qubit $|\psi\rangle$ from A to B. This is not true, since A's $|q_1\rangle$ qubit collapsed onto either a $|0\rangle$ or $|1\rangle$ upon measurement during the process of teleportation (and so we are left with exactly one copy $|\psi\rangle$ in B's possession).

3.6 Superdense Coding

In quantum teleportation, we used two classical bits to send a qubit. Superdense coding is simply the reverse: a method to send two bits of information using one qubit.

Theorem 3.6 (Superdense Coding)

Given that A wants to send B two bits of information with just a single qubit, it is possible to do so under the following assumptions:

1. A Bell state has been generated with each qubit given to A and B. This can be done beforehand by A, B, or even a third party C.
2. There exists a channel where A can send a qubit to B.

The entire circuit is represented by the diagram below:

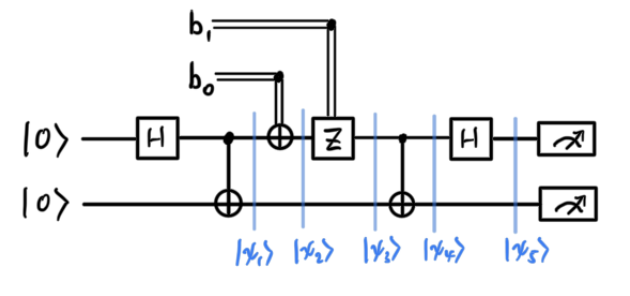


Figure 5: Superdense coding circuit, with A having the first qubit with the 2 classical bits and B having the second qubit. Once A manipulates the first qubit, A sends the first qubit to B, who then measures the first and second qubits to get the 2 classical bits.

The following steps are taken.

1. Before the experiment, we take the $|00\rangle$ qubit and entangle it, bringing us to $|\psi_1\rangle$.
$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\big(\,|00\rangle + |11\rangle\,\big) \tag{130}$$
2. Then we do a control NOT with a classical control bit b_0.
$$|\psi_2\rangle = \frac{1}{\sqrt{2}}\big(\,|0b_0\rangle + |1, 1 \oplus b_0\rangle\,\big) \tag{131}$$
3. Then we do a control Z with a classical control bit b_1
$$|\psi_3\rangle = \frac{1}{\sqrt{2}}\big(\,|0b_0\rangle + (-1)^{b_1}\,|1, 1 \oplus b_0\rangle\,\big) \tag{132}$$
4. Then we do a quantum control NOT with the control qubit as the first qubit that A has, which unentangles the system!
$$|\psi_4\rangle = \frac{1}{\sqrt{2}}\big(\,|0b_0\rangle + (-1)^{b}\,|1b_1\rangle\,\big) = \left(\frac{|0\rangle + (-1)^{b_1}\,|1\rangle}{\sqrt{2}}\right)|b_0\rangle \tag{133}$$

5. Then we do a Hadamard on the first qubit unentangling it.

$$|\psi_5\rangle = (H \otimes I)\,|\psi_4\rangle = |b_1 b_0\rangle \tag{134}$$

6. Now the first qubit, in possession by A, is sent to B, which is now in state $|b_1 b_0\rangle$.

Example 3.5 (Superdense Coding with Two Classical 0 Bits)

Since it is a bit more tedious to prove, we will show for the $(b_0, b_1) = (0, 0)$ case and show for the $(b_0, b_1) = (0, 1)$ case in a future example.

1. Before the experiment, we take the $|00\rangle$ qubit and entangle it, bringing us to $|\psi_1\rangle$.

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\big(\,|00\rangle + |11\rangle\,\big) \tag{135}$$

2. Then we do a control NOT with a classical control bit $b_0 = 0$. Nothing changes.

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}\big(\,|00\rangle + |11\rangle\,\big) \tag{136}$$

3. Then we do a control Z with a classical control bit $b_1 = 0$. Nothing changes.

$$|\psi_3\rangle = \frac{1}{\sqrt{2}}\big(\,|00\rangle + |11\rangle\,\big) \tag{137}$$

4. Then we do a quantum control NOT with the control qubit as the first qubit that A has.

$$|\psi_4\rangle = \frac{1}{\sqrt{2}}\big(\,|00\rangle + |10\rangle\,\big) = \frac{1}{\sqrt{2}}\big(\,|0\rangle + |1\rangle\,\big)\,|0\rangle \tag{138}$$

5. Then we do a Hadamard on the first qubit unentangling it.

$$|\psi_5\rangle = (H \otimes I)\,|\psi_4\rangle = |00\rangle \tag{139}$$

6. Now the first qubit, in possession by A, is sent to B.
7. The final qubit is of state $|00\rangle$, which is guaranteed to have measurement $(0, 0) = (b_0, b_1)$ as A had, and so B measures the first and second qubits to get the 2 classical bits.

Example 3.6 (Superdense Coding with Two Classical 1 Bits)

As promised, we outline the steps for when A has $(b_0, b_1) = (1, 1)$.

1. Before the experiment, we take the $|00\rangle$ qubit and entangle it, bringing us to $|\psi_1\rangle$.

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\big(\,|00\rangle + |11\rangle\,\big) \tag{140}$$

2. Then we do a control NOT with a classical control bit $b_0 = 1$. The first qubit is flipped.

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}\big(\,|10\rangle + |01\rangle\,\big) \tag{141}$$

3. Then we do a control Z with a classical control bit $b_1 = 1$. The $|1\rangle$ component of the first qubit

changes sign from + to −.

$$|\psi_3\rangle = \frac{1}{\sqrt{2}}\big(-|10\rangle + |01\rangle \big) \tag{142}$$

4. Then we do a quantum control NOT with the control qubit as the first qubit that A has.

$$|\psi_4\rangle = \frac{1}{\sqrt{2}}\big(-|11\rangle + |01\rangle \big) = \frac{1}{\sqrt{2}}\big(|0\rangle - |1\rangle \big) |1\rangle \tag{143}$$

5. Then we do a Hadamard on the first qubit unentangling it.

$$|\psi_5\rangle = (H \otimes I) |\psi_4\rangle = |11\rangle \tag{144}$$

6. Now the first qubit, in possession by A, is sent to B.

7. The final qubit is of state $|11\rangle$, which is guaranteed to have measurement $(1, 1) = (b_0, b_1)$ as A had, and so B measures the first and second qubits to get the 2 classical bits.

3.7 Quantum Parallelism

There are two limitations of quantum computing. The first one is that the gates must be $n \times n$ and therefore the number of qubits coming in must equal the number of qubits coming out. This is not too bad, since we can add, remove, or measure in some clever way. The bigger limitation is that quantum gates must be reversible, and therefore they cannot simulate functions that are not invertible. For example, say the function $f : \{0, 1\} \longrightarrow \{0, 1\}$ is defined by $f(0) = f(1) = 1$. This is not invertible, and so we cannot simulate this function with a quantum gate. What we can do is simply extend the codomain so that no outputs overlap with each other. That is, we can simply say $f(0) = 01$ and $f(1) = 11$, setting the first bit to be a dummy variable and the second bit to be the actual output. But now we've broken the first rule since the number of qubits coming out is not the same as the number of qubits coming in. Therefore, we simply increase the dimensionality of the domain as well. Therefore, we can say that $f(00) = 01$ and $f(10) = 11$, where the second bit of the input is now a dummy variable. Problem solved. It is not too hard to see the general case.

Lemma 3.2 (Reversible Computing of a Classical Function)

Given a classical function $f : \{0, 1\}^n \longrightarrow \{0, 1\}^m$, we can simulate this function with a reversible quantum gate $U_f : \{0, 1\}^{n+m} \longrightarrow \{0, 1\}^{n+m}$, where n is the number of input bits and m is the number of output bits.

Proof.

We don't give a formal proof here, but consider the worst case scenario where $f = 0^m$ always. Since all the outputs are the same, the inputs must be distinct, so we need at least n dimensions to store the keys of the inputs. The actual output size is m, so we also need an extra m dimensions to store the outputs, leading to $n + m$.

Therefore, we have the set of inputs $x \in \{0, 1\}^n$ and the set of outputs $\{f(x)\}$. It is conventional to do the following:

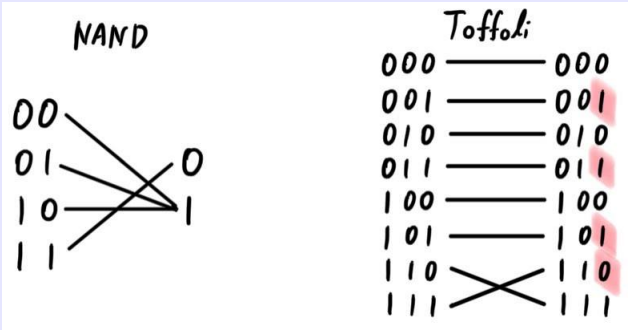
1. Since the inputs must be of dimension $n + m$, we just write the first n bits as the actual inputs and the last m bits as dummy variables, call it y, which can be set to all 0 or all 1 for simplicity.

2. Since the outputs is of dimension $n + m$, the first n bits will be set to the dummy variables, which are set to the input values, and the last m bits will be the actual outputs.

For now, let's focus on when $m = 1$.

Example 3.7 (Toffoli Gate as extension of NAND Gate)

The Toffoli gate can simulate the NAND gate as its extension. To see how, the input is of form $\mathbf{x}y = \mathbf{x_1 x_2 y} = x_1 x_2 1$, where we have set y to all 1. Then looking at the output of the Toffoli gate, we can see that in its respective outputs, the first two bits are simply the inputs and the third bit is the NAND of the first two bits, the actual output.



More formally, we can write

$$\text{NAND}(xy) = \delta_3 \circ \text{Toffoli}(xy1) \tag{145}$$

where $\delta_3 : xyz \mapsto z$. Since the Toffoli gate allows us to simulate the universal NAND gate, it becomes possible to simulate all other elements in a classical circuit and thus an arbitrary classical circuit can be simulated by an equivalent reversible circuit.

Definition 3.5 (Reversible Extension of Functions)

In general, given a function $f : \{0,1\}^n \to \{0,1\}$, we can set the dummy variable $y = 0$ always and construct a reversible extension U_f defined

$$U_f : \{0,1\}^{n+1} \longrightarrow \{0,1\}^{n+1},\ U_f : |x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle. \tag{146}$$

and since $y = 0$, $y \oplus f(x) = f(x)$.

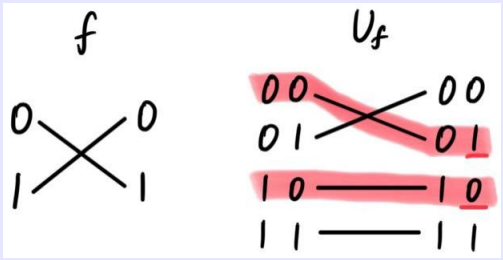
Constructing this reversible extension of classical functions gives us the foundation to work with **quantum parallelism**, which is a fundamental feature of many quantum algorithms that, heuristically, allows quantum computers to evaluate a function $f(x)$ for many different values of x simultaneously. More specifically, by inputting a superposition of all possible inputs into a reversible extension of a function, we can obtain a superposition of all possible outputs.

Example 3.8 (Function of 1-Bit Input)

Suppose $f : \{0,1\} \longrightarrow \{0,1\}$ is a function. To evaluate f on all possible bits, we need to call f 2 times: $f(0), f(1)$. We can create a reversible extension of this function f by extending the domain and codomain to $\{0,1\}^2$ to construct

$$U_f : \{0,1\}^2 \longrightarrow \{0,1\}^2,\ U_f : |x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle \tag{147}$$

With the construction of U_f, all we need to do is pay attention to all the outputs where the input has $y = 0$, since $U_f(|x\rangle|0\rangle) = |x\rangle\,|f(x)\rangle$.



By setting $|y\rangle = |0\rangle$ and with it, (the important part) inputting in a superposition of all possible inputs (which we can just get by doing a Hadamard) $|x\rangle = H|0\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$, we can obtain a superposition of all possible outputs.

$$\begin{aligned} U(|x\rangle \otimes |y\rangle) &= U\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes |0\rangle\right) \\ &= U\left(\frac{|00\rangle + |10\rangle}{\sqrt{2}}\right) \\ &= \frac{U|00\rangle + U|10\rangle}{\sqrt{2}} \\ &= \frac{|0\rangle|f(0)\rangle + |1\rangle|f(1)\rangle}{\sqrt{2}} \end{aligned}$$

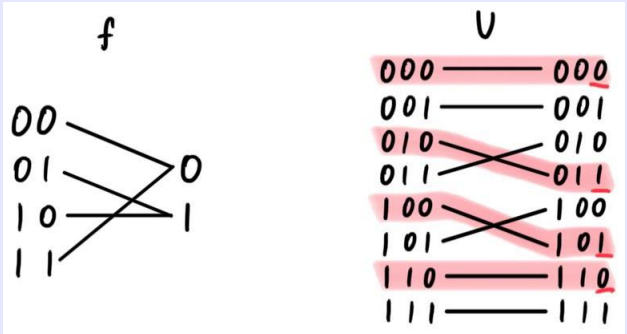
This output state is very interesting because the different terms contain information about both $f(0)$ and $f(1)$. It is almost as if we evaluated $f(x)$ for two values of x simultaneously. Note that unlike classical parallelism, where multiple circuits each built to compute $f(x)$ are executed simultaneously, here a single $f(x)$ circuit is employed to evaluate the function for multiple values of x simultaneously.

Example 3.9 (Function of 2-Bit Input)

Suppose $f : \{0,1\}^2 \longrightarrow \{0,1\}$ is a function. To evaluate f on all four permutations of two bits, we need to call f 4 times: $f(00), f(01), f(10), f(11)$. We can create a reversible extension of this function f by extending the domain and codomain to $\{0,1\}^3$ to construct

$$U_f : \{0,1\}^3 \longrightarrow \{0,1\}^3, \;\; U_f : |x_1x_2\rangle|y\rangle \mapsto |x_1x_2\rangle|y \oplus f(x_1x_2)\rangle \tag{148}$$

With the construction of U_f shown below, all we need to do is pay attention to all the outputs where the input has $y = 0$, since $U_f(|x_1x_2\rangle|0\rangle) = |x_1x_2\rangle|f(x)\rangle$.



Again, we set $|y\rangle = 0$ and now we want a superposition of all possible two-qubit inputs. What better

way to do this than to apply the Hadamard transform $H^{\otimes 2}$ to the first two qubits?

$$U(|x\rangle \otimes |y\rangle) = U\big(H^{\otimes 2} |00\rangle \otimes |0\rangle\big) \tag{149}$$

$$= \dots \tag{150}$$

$$= \frac{|00\rangle |f(00)\rangle + |01\rangle |f(01)\rangle + |10\rangle |f(10)\rangle + |11\rangle f(|11\rangle)}{2} \tag{151}$$

Again, this output state is interesting because it contains information about all its input values. It is almost as if we evaluated $f(x)$ simultaneously.

If you can simulate a NAND gate with a Toffoli, and NAND is the universal gate for classical computing, then can't we construct cloning gates (like XOR)? The answer is no, since we can only do it for classical bits, not quantum bits. Let's generalize this procedure for N-bit inputs.

Theorem 3.7 (Quantum Parallelism of N-bit Input)

Given some function $f : \{0,1\}^n \longrightarrow \{0,1\}$, we can take the state $|0\rangle^{\otimes(n+1)}$ and apply the Hadamard transformation $H^{\otimes n}$ on the first n qubits to prepare the $(n+1)$-qubit state

$$|x_1 \dots x_n\rangle \otimes |y\rangle = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)^{\otimes n} \otimes |0\rangle \tag{152}$$

Then, we put it through the (reversible) quantum circuit $U_f : \{0,1\}^{n+1} \longrightarrow \{0,1\}^{n+1}$ constructed as an extension of the classical f defined

$$U_f |x_1 x_2 \dots x_n\rangle |y\rangle \mapsto |x_1 x_2 \dots x_n\rangle |y \oplus f(x_1 \dots x_n)\rangle \tag{153}$$

which gives

$$U_f\big(H^{\otimes n} |0\rangle^{\otimes n} \otimes |0\rangle\big) = U_f\left(\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)^{\otimes n} \otimes |0\rangle\right) = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle |f(x)\rangle \tag{154}$$

This output state contains information about all of the possible values $f(x)$. But the question still remains what to do with this output

$$\frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle |f(x)\rangle \tag{155}$$

While this one state contains all the information defining the function f on the one hand, it is still in superposition that will collapse onto *one* measurement outcome $|x\rangle \otimes |f(x)\rangle$. Therefore, quantum computation requires something more than just quantum parallelism to be useful; it requires the ability to *extract* information about more than one value of $f(x)$ from superposition states. This is where the Deutsch algorithm comes in.

4 Quantum Fourier Transform

4.1 Deutsch Algorithm

With quantum parallelism, we've evaluated a superposition of the inputs to get a superposition of the outputs. Unfortunately, we can't just extract all information from this output, but there are some *global properties* of the function that we can deduce, with *interference*. What do we mean by global properties? Let's look at an example below.

Example 4.1 (1 Bit Function)

Given a function $f : \{0,1\} \longrightarrow \{0,1\}$, we can determine the following equivalent properties:

1. whether the f is constant or balanced, i.e. whether it is constantly 0 or 1.
2. the value of $f(0) \oplus f(1)$
3. whether f is surjective or not.

It is easy to see how knowing one property allows you to know other properties.

To get these global properties, it usually requires us to run a classical circuit over all inputs of f, which may not be computationally feasible. Finding the surjectivity of a 1-bit input function above will require 2 forward passes, but we claim that we can do it on a quantum circuit with just one forward pass. To do this, it requires a bit of clever thinking, where the dummy variable $|y\rangle$ is now also set to a Bell state.

Theorem 4.1 (Deutsch Algorithm for 1-Bit Input Function)

Let Alice and Bob be isolated except for a channel where they may send one qubit to each other. Given that Bob has a one-bit input function $f : \{0,1\} \longrightarrow \{0,1\}$, it is possible for Alice to determine the value of $f(0) \oplus f(1)$ with just one forward pass of $f(x)$ on a quantum circuit without knowing Bob's function. The circuit is shown first for clarity:

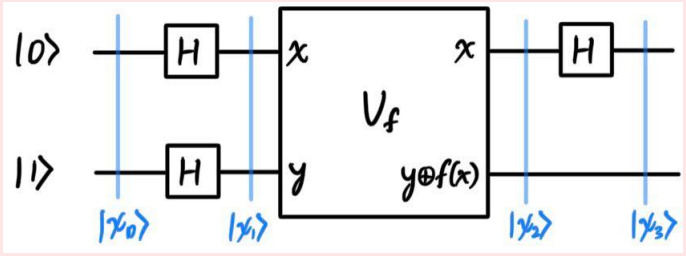


Figure 6: Deutsch Algorithm for 1-bit input function.

Like we have always done, we can take the reversible extension $U_f : |x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle$. The following steps are taken.

1. Alice prepares a bell state by taking a Hadamard on each input qubit.
$$|\psi_1\rangle = H^{\otimes 2} |01\rangle = \left|\Phi^+\right\rangle \otimes \left|\Phi^-\right\rangle \tag{156}$$
2. Alice sends $|\psi_1\rangle$ to Bob, and Bob now passes it through U_f, but before this, observe that for any arbitrary input x, we have
$$U_f\left(|x\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \right) = \frac{U_f(|x\rangle |0\rangle) - U_f(|x\rangle |1\rangle)}{\sqrt{2}} \tag{157}$$
$$= \frac{|x\rangle \otimes |0 \oplus f(x)\rangle - |x\rangle \otimes |1 \oplus f(x)\rangle}{\sqrt{2}} \tag{158}$$
$$= \frac{|x\rangle \otimes \big(|0 \oplus f(x)\rangle - |1 \oplus f(x)\rangle \big)}{\sqrt{2}} \tag{159}$$
$$= \begin{cases} \frac{|x\rangle \otimes (|0\rangle - |1\rangle)}{\sqrt{2}} & \text{if } f(0) = f(1), \\ \frac{-|x\rangle \otimes (|0\rangle - |1\rangle)}{\sqrt{2}} & \text{if } f(0) \neq f(1) \end{cases} \tag{160}$$
$$= \frac{(-1)^{f(x)} |x\rangle \big(|0\rangle - |1\rangle \big)}{\sqrt{2}} \tag{161}$$

Therefore, for $|x\rangle = |\Phi^+\rangle$, we have

$$|\psi_2\rangle = U_f |\psi_1\rangle = U_f\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \tag{162}$$

$$= \frac{1}{\sqrt{2}} U_f\left(|0\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) + \frac{1}{\sqrt{2}} U_f\left(|1\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \tag{163}$$

$$= \frac{(-1)^{f(0)} |0\rangle (|0\rangle - |1\rangle)}{\sqrt{2}\cdot\sqrt{2}} + \frac{(-1)^{f(1)} |1\rangle (|0\rangle - |1\rangle)}{\sqrt{2}\cdot\sqrt{2}} \tag{164}$$

$$= \begin{cases} \pm\left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right) & \text{if } f(0) = f(1), \\ \pm\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right) & \text{if } f(0) \neq f(1) \end{cases} \tag{165}$$

3. Bob sends $|\psi_2\rangle$ back to Alice, who now applies the Hadamard gate on the first qubit now gives us

$$|\psi_3\rangle = (H \otimes I) |\psi_2\rangle = \begin{cases} \pm |0\rangle \left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right) & \text{if } f(0) = f(1), \\ \pm |1\rangle \left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right) & \text{if } f(0) \neq f(1) \end{cases} \tag{166}$$

$$= \pm |f(0) + f(1)\rangle \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \tag{167}$$

4. Therefore, by measuring the first qubit, we can determine $f(0) \oplus f(1)$, i.e., whether $f(0) = f(1)$ or $f(0) \neq f(1)$.

4.2 The Deutsch-Jozsa Algorithm

This type of problem where we must efficiently guess this property of a function is called Deutch's problem.

Definition 4.1 (Deutch's Problem)

Alice selects a number x from 0 to $2^n - 1$, which can be represented as an element in $\{0,1\}^n$ (isomorphic?) and mails it in a letter to Bob. Bob calculates some function $f : \{0,1\}^n \longrightarrow \{0,1\}$ and replies with the result, which is either 0 or 1. Bob promises to use a function f which is one of two kinds:

1. $f(x)$ is constant for all values of x, or
2. $f(x)$ is balanced, meaning that it outputs 1 for exactly half of the possible x and 0 for the other half (note that this does *not* mean that f outputs 0 and 1 probabilistically; f is completely deterministic).

However, this problem is not known to have any applications, and probabilistic computation can be similarly used to solve this problem with a high (but not certain) degree of accuracy.

Again, in the classical case, Alice may send Bob one value of x in each letter. At worst, she will need to query Bob at least $2^{n-1} + 1$ times (half of possible inputs, plus one) and therefore the best deterministic classical algorithm she can use therefore requires $2^{n-1} + 1$ queries (i.e. a computational complexity of $O(2^n)$). However, we claim that if Bob and Alice were able to exchange qubits instead of classical bits, and Bob agreed to calculate $f(x)$ using a unitary transformation U_f, then Alice can achieve her goal of just *one* correspondence with Bob. Before we start, let's introduce a lemma.

Lemma 4.1 ()

Given a state $|x\rangle$, the Hadamard gate H acts on it as

$$H|x\rangle = \sum_{z\in\{0,1\}} \frac{(-1)^{x\cdot z}|z\rangle}{\sqrt{2}} \tag{168}$$

and for general state $|\mathbf{x}\rangle = |x_1x_2\dots x_n\rangle$, we have

$$H^{\otimes n}|\mathbf{x}\rangle = \sum_{\mathbf{z}\in\{0,1\}^n} \frac{(-1)^{\mathbf{x}\cdot\mathbf{z}}|\mathbf{z}\rangle}{\sqrt{2^n}} \tag{169}$$

where $\mathbf{x}\cdot\mathbf{z}$ is the bitwise dot product of $\mathbf{x},\mathbf{z}\in\{0,1\}^n$ modulo 2.

Proof.

We can see that for $|x\rangle = |0\rangle$ or $|1\rangle$,

$$H|x\rangle = \sum_{z\in\{0,1\}} \frac{(-1)^{xz}|z\rangle}{\sqrt{2}} = \frac{|0\rangle + (-1)^x|1\rangle}{\sqrt{2}} \tag{170}$$

Now given $|x\rangle = |x_1x_2\dots x_n\rangle$, we can take the tensor products of these terms.

$$\begin{aligned}
H^{\otimes n}|\mathbf{x}\rangle &= H|x_1\rangle \otimes H|x_2\rangle \otimes \dots \otimes H|x_n\rangle \\
&= \left(\sum_{z_1\in\{0,1\}} \frac{1}{\sqrt{2}}(-1)^{x_1z_1}|z_1\rangle\right) \otimes \dots \otimes \left(\sum_{z_n\in\{0,1\}} \frac{(-1)^{x_nz_n}|z_n\rangle}{\sqrt{2}}\right) \\
&= \sum_{\mathbf{z}\in\{0,1\}^n} \frac{(-1)^{x_1z_1}|z_1\rangle \otimes \dots \otimes (-1)^{x_nz_n}|z_n\rangle}{\sqrt{2^n}} \\
&= \sum_{\mathbf{z}\in\{0,1\}^n} \frac{(-1)^{x_1z_1+\dots+x_nz_n}|z_1\dots z_n\rangle}{\sqrt{2^n}} \\
&= \sum_{\mathbf{z}\in\{0,1\}^n} \frac{(-1)^{\mathbf{x}\cdot\mathbf{z}}|\mathbf{z}\rangle}{\sqrt{2^n}}
\end{aligned}$$

Theorem 4.2 (Deutsch Jozsa Algorithm)

Let Alice and Bob be isolated except for a channel where they may send qubits to each other. Alice can solve Deutch's problem with a single pass on a quantum circuit. The circuit is shown first for clarity:

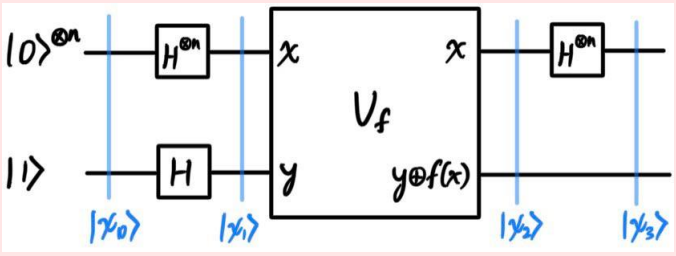


Figure 7: Deutsch-Jozsa Algorithm.

Bob constructs the reversible extension $U_f : |x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle$. The following steps are taken.

1. Alice prepares a state $|\psi_1\rangle = H^{\otimes n} |0\rangle^{\otimes n} \otimes |1\rangle$ and sends it to Bob.

$$|\psi_1\rangle = H^{\otimes(n+1)}(|0\rangle^{\otimes n} \otimes |1\rangle) = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{171}$$

2. Bob applies U_f to $|\psi_1\rangle$ to get

$$|\psi_2\rangle = U_f |\psi_1\rangle = \sum_{x\in\{0,1\}^n} \frac{(-1)^f(x)\,|x\rangle}{\sqrt{2^n}} \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{172}$$

3. Bob sends $|\psi_2\rangle$ back to Alice, who evaluates the Hadamard on the first n qubits. We can use the previous lemma to get

$$|\psi_3\rangle = H^{\otimes n} |\psi_2\rangle = \sum_{\mathbf{z}} \sum_{\mathbf{x}} \frac{(-1)^{\mathbf{x}\cdot\mathbf{z}+f(\mathbf{x})}\,|\mathbf{z}\rangle}{2^n} \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{173}$$

4. Alice now measures the query register (i.e., the first n qubits). The complete expansion is too long to write out, but we can focus on the amplitude for the state $|\mathbf{z}\rangle = |0\rangle^{\otimes n}$.

$$\sum_{z\in\{0,1\}^n} \sum_{x\in\{0,1\}^n} \frac{(-1)^{x\cdot z+f(x)}}{2^n}|z\rangle = \left(\sum_{x\in\{0,1\}^n} \frac{(-1)^{f(x)}}{2^n}\right) |0\dots0\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} + \dots \tag{174}$$

Note that the amplitude for the state $|0\rangle^{\otimes n}$ (i.e., when $|z\rangle = |0\dots0\rangle$) is

$$\sum_{x\in\{0,1\}^n} \frac{(-1)^{f(x)}}{2^n} \tag{175}$$

There are two scenarios:

(a) If $f(x)$ is constant for all values of x, then this amplitude would be

$$\sum_{x\in\{0,1\}^n} \frac{(-1)^{f(x)}}{\sqrt{2^n}} = \sum_{x\in\{0,1\}^n} \frac{1}{2^n} = 1 \text{ if } f(x) = 0,$$
$$\sum_{x\in\{0,1\}^n} \frac{(-1)^{f(x)}}{\sqrt{2^n}} = \sum_{x\in\{0,1\}^n} \frac{-1}{2^n} = -1 \text{ if } f(x) = 1.$$

Since $|\psi_3\rangle$ must be a unit vector, this implies that all the other amplitudes are 0 and therefore $|\psi_3\rangle = |0\rangle^{\otimes n}$.

(b) If $f(x)$ is balanced, then the number of negative terms and positive terms will cancel each other out, and so

$$\sum_{x\in\{0,1\}^n} \frac{(-1)^{f(x)}}{2^n} = 2^{n-1}\cdot 1 + 2^{n-1}\cdot(-1) = 0,$$

meaning that the amplitude of $|0\rangle^{\otimes n}$ is 0.

Therefore, by measuring the query register, if everything pops up to **1**, then it is constant and if everything collapses to **0**, then it is balanced.

Proof.

We've left out a lot of computation here for simplicity. Here we show the full derivations.

1. Trivial.
2. Again, to calculate $|\psi_2\rangle$, we use the fact (which does not matter how many bits x is)

$$U_f|x\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} = (-1)^{f(x)}|x\rangle \frac{|1\rangle - |0\rangle}{\sqrt{2}},$$

and sum it over all possible permutations of $x \in \{0,1\}^n$ to get

$$\begin{aligned} U_f|\psi_1\rangle &= U_f\left(\sum_x \frac{1}{\sqrt{2}}|x\rangle\right) \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \\ &= U_f \sum_x \left(\frac{1}{\sqrt{2}}|x\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \\ &= \sum_x \frac{1}{\sqrt{2}} U_f\left(|x\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \\ &= \sum_x (-1)^{f(x)} \frac{|x\rangle}{\sqrt{2^n}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \end{aligned}$$

3. $|\psi_3\rangle$ is calculated with the following steps.

$$\begin{aligned} |\psi_3\rangle = (H^{\otimes n}\otimes I)|\psi_2\rangle &= H^{\otimes n}\left(\sum_{x\in\{0,1\}^n} (-1)^{f(x)}\frac{|x\rangle}{\sqrt{2^n}}\right) \otimes I\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \\ &= \left(\sum_{x\in\{0,1\}^n} \frac{(-1)^{f(x)}}{\sqrt{2^n}} H^{\otimes n}|x\rangle\right) \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \\ &= \sum_{x\in\{0,1\}^n} \frac{(-1)^{f(x)}}{\sqrt{2^n}} \left(\frac{1}{\sqrt{2^n}} \sum_{z\in\{0,1\}^n} (-1)^{x\cdot z}|z\rangle\right) \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \\ &= \sum_{x\in\{0,1\}^n} \sum_{z\in\{0,1\}^n} \frac{1}{2^n}(-1)^{x\cdot z + f(x)}|z\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \\ &= \sum_{z\in\{0,1\}^n} \sum_{x\in\{0,1\}^n} \frac{1}{2^n}(-1)^{x\cdot z + f(x)}|z\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \end{aligned}$$

Remember that quantum computation is superior to classical computation if *both* of the following requirements are met:

1. We can utilize the non-binary superpositions of qubits to calculate more efficiently using quantum parallelism.
2. We have some method to *extract* information from the output qubit(s) using measurements. The answers are all there in the qubit state, but they are hidden: measuring it would cause it to collapse onto a string of $|0\rangle$s and $|1\rangle$s, and so a creative method of gaining information is needed.

Theorem 4.3 (Quantum Fourier Transform)

The quantum Fourier transform is a unitary transformation that maps a state $|x\rangle$ to a state $|\tilde{x}\rangle$ according to the rule

$$QFT|\mathbf{x}\rangle = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{\frac{2\pi i \mathbf{x}\mathbf{k}}{2^n}} |\mathbf{k}\rangle \tag{176}$$

$$= \frac{|0\rangle + e^{2\pi i(0.x_n)}|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle + e^{2\pi i(0.x_{n-1}x_n)}|1\rangle}{\sqrt{2}} \otimes \cdots \otimes \frac{|0\rangle + e^{2\pi i(0.x_1x_2\cdots x_n)}|1\rangle}{\sqrt{2}} \quad . \tag{177}$$

Proof.

$$\frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{\frac{2\pi i \mathbf{x}\mathbf{k}}{2^n}} |\mathbf{k}\rangle = \frac{1}{\sqrt{2^n}} \sum_{k_1 \in \{0,1\}} \cdots \sum_{k_n \in \{0,1\}} \exp\left(2\pi i \mathbf{x} \Big[\sum_{l=1}^{n} k_l 2^{-l}\Big]\right) |k_1 \dots k_n\rangle \tag{178}$$

$$= \frac{1}{\sqrt{2^n}} \sum_{k_1 \in \{0,1\}} \cdots \sum_{k_n \in \{0,1\}} \bigotimes_{l=1}^{n} e^{2\pi i \mathbf{x} k_l 2^{-l}} |k_l\rangle \tag{179}$$

$$= \frac{1}{\sqrt{2^n}} \bigotimes_{l=1}^{n} \sum_{k_l \in \{0,1\}} e^{2\pi i \mathbf{x} k_l 2^{-l}} |k_l\rangle \tag{180}$$

$$= \frac{1}{\sqrt{2^n}} \bigotimes_{l=1}^{n} \left(|0\rangle + e^{2\pi i \mathbf{x} 2^{-l}} |1\rangle\right) \tag{181}$$

$$= \frac{1}{\sqrt{2^n}} \bigotimes_{l=1}^{n} \left(|0\rangle + e^{2\pi i (0.x_l)} |1\rangle\right) \tag{182}$$

$$= \frac{|0\rangle + e^{2\pi i(0.x_n)}|1\rangle}{\sqrt{2}} \otimes \cdots \otimes \frac{|0\rangle + e^{2\pi i(0.x_1x_2\cdots x_n)}|1\rangle}{\sqrt{2}} \tag{183}$$

where we note that $e^{2\pi i \mathbf{x}}$ is periodic, and so the terms to the left of the dot in the binary expansion can be ignored.

4.3 Phase Estimation
4.4 Order Finding
4.5 Factoring
4.6 Period Finding
4.7 Discrete Logarithms
4.8 Hidden Subgroup Problems
4.9 Shor's Algorithm

5 Quantum Error Correction

6 Quantum Simulation

Neilsen and Chuang and of chapter 4.

7 Quantum Search Algorithms

8 Class

Definition 8.1 (Separable)

Density matrix ρ is **separable** iff

$$\rho = \sum_i p_i \rho_{i,A} \otimes \rho_{i,B} \tag{184}$$

else it is entangled.

Definition 8.2 (Werner State)

A Werner state is a state of the form

$$\rho = \eta \left|\Phi^+\right\rangle \left\langle\Phi^+\right| + (1-\eta)\frac{I}{4} \tag{185}$$

where $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. You have some entangled state, and you are mixing into a it a maximally mixed state.

Now we can take some density matrix $\bar{\rho}_1$ and decompose it as simple the sum of separable density matrices. Note that these are not technically density matrices since they must be normalized, so there is an extract factor of $1/4, 1/2, 1/2$ on $\bar{\rho}_1, \bar{\rho}_2, \bar{\rho}_3$ respectively.

$$\bar{\rho}_1 = \rho \left|++\right\rangle \left\langle++\right| + \rho \left|--\right\rangle \left\langle--\right| + \rho \left|+i,-i\right\rangle \left\langle+i,-i\right| + \rho \left|-i,+i\right\rangle \left\langle-i,+i\right| \tag{186}$$

$$= \begin{pmatrix} 1 & & & 1 \\ & 1 & & \\ & & 1 & \\ 1 & & & 1 \end{pmatrix} \tag{187}$$

Note also that

$$\bar{\rho}_2 = |00\rangle \langle 00| + |11\rangle \langle 11| \tag{188}$$

and

$$\bar{\rho}_3 = |01\rangle \langle 01| + |10\rangle \langle 10| \tag{189}$$

Therefore after some calculations, we have another decomposition of a density matrix

$$\rho(\eta) = \frac{\eta}{2}\bar{\rho}_1 + \frac{1-\eta}{4}\bar{\rho}_2 + \left(\frac{1-\eta}{4} - \frac{\eta}{2}\right)\bar{\rho}_3 \tag{190}$$

$$= \frac{\eta}{2}4 \cdot \frac{\bar{\rho}_1}{4} + \frac{1-\eta}{4}2 \cdot \frac{\bar{\rho}_2}{2} + \frac{1-\eta}{4}2 \cdot \frac{\bar{\rho}_3}{2} \tag{191}$$

Just as a sanity check, the coefficients must sum to 1. This is true since

$$\frac{\eta}{2} \cdot 4 + \frac{1-\eta}{4} \cdot 2 + \frac{1-3\eta}{4} \cdot 2 = 1 \tag{192}$$

This is valid if all the coefficients are positive, which means that $\eta \leq 1/3$. Therefore, if $\eta \leq 1/3$, then $\rho(\eta)$ is separable. It turns out that as we decrease η from 1 down to 0, it starts off entangled and then at $\eta = 1/3$ it vanishes. This is very surprising.

Question 8.1 ()

Is this because of matrix eigenbases being discontinuous over small perturbations of the elements?

Definition 8.3 (Schmidt Decomposition)

We can decompose a n qubit vector into

$$|\psi\rangle = \sum_{jk} C_{jk} |j\rangle |k\rangle \tag{193}$$

we take the SVD of $C = UDV$ and then we can write the above as

$$|\psi\rangle = \sum_{ijk} U_{ji} D_{ii} V_{ik} |j\rangle |k\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle \tag{194}$$

where $\lambda_i = D_{ii}$ are the **Schmidt coefficients**, $|i_A\rangle = U_{ji} |j\rangle$, and $|i_B\rangle = V_{ik} |k\rangle$.

Lemma 8.1 ()

The density matrix ρ is a product state iff its Schmidt number (the number of non-zero Schmidt coefficients) is 1. Therefore, the local density matrices

$$\rho_A = \sum_i \lambda_i^2 |i_A\rangle \langle i_A| \tag{195}$$

which implies non zero eigenvalues are the same for ρ_A, ρ_B. It is also the case that the purity matches $\text{Tr}(\rho_A^2) = \text{Tr}(\rho_B^2)$ and the von Neumann entropy is the same $S(\rho_A) = S(\rho_B)$. Also

$$U_A \otimes U_b |\psi\rangle = \sum_i \lambda_i U_A |i_A\rangle \otimes U_B |i_B\rangle \tag{196}$$

Recall the definition of Shannon entropy.

Definition 8.4 (Shannon Entropy)

The Shannon entropy of a probability distribution $p(x)$ is

$$H(p) = -\sum_x p(x) \log p(x) \tag{197}$$

where we take the convention that $0 \log 0 = 0$.

Theorem 8.1 (Subsystem Entropy)

Classicaly, the subsystem has a smaller entropy than the whole system.

$$H(x) \leq H(x, y) \tag{198}$$

Definition 8.5 (Von Neumann Entropy)

The von Neumann entropy of a density matrix ρ is

$$S(\rho) = -\operatorname{Tr}(\rho \log \rho) \tag{199}$$

Quantum mechanically, this is not necessarily true, but this is a necessary (not sufficient) condition for separability.

$$S(A) \text{ or } S(B) \leq S(A, B) \tag{200}$$

This means that subsystems may exhibit *more* entropy than the whole system. In fact, it turns out that

$$S(A) \geq S(A, B) \tag{201}$$

is a sufficient condition for entanglement.

Now let's apply this to the Weiner State, where I use $\log_2$

$$S(A) = 1 \tag{202}$$

$$S(A, B) = H\left(\frac{1-\eta}{4}, \frac{1-\eta}{4}, \frac{1-\eta}{4}, \frac{1+3\eta}{4}\right) \tag{203}$$

If we graph this, we will see a monotonically decreasing function of η, and $\eta = 0.7476$ is the critical point where it becomes entangled. Now let's talk about majorization.

Definition 8.6 (Majorization)

We say x is majorized by y, or $x \prec y$, if

$$\sum_{i=1}^{k} x_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow} \tag{204}$$

Theorem 8.2 (Equivalent Property for Majorization)

$x \prec y$ iff $x = Dy$, where D is a doubly stochastic matrix.

Look at Neilsen and Kempe 2000, which states that separable states are more disordered globally than locally.

Theorem 8.3 ()

If ρ_{AB} is separable, then $\lambda(\rho_{AB}) \prec \lambda(\rho_B)$.