Quantum Computation Theory

Arpon Basu

Last updated March 3, 2025

Contents

§1. Basics of Quantum Computation

Conventions

We order multi-qubit states lexicographically while expressing them as vectors, where $|0\rangle \prec |1\rangle$. Thus, for example, $\begin{bmatrix}0.7 & 0.5 & -0.5 & 0.1\end{bmatrix}^{\mathsf{T}}$ stands for $0.7|00\rangle + 0.5|01\rangle - 0.5|10\rangle + 0.1|11\rangle$.
Let M be a $m \times n$ complex matrix. We denote by M^{T} the transpose of M, M^* the conjugate of M, and $M^\dagger$ the conjugate transpose of M. Thus $M^* \in \mathbb{C}^{m\times n}, M^{\mathsf{T}}, M^\dagger \in \mathbb{C}^{n\times m}$.
Let V be a vector space, and let $M : V \to V$ be a linear map. Let $\mathcal{V} := \{v_i\}$ be a basis for V. Then expressing M in the basis $\mathcal{V}$ means creating a $\dim(V) \times \dim(V)$ matrix such that $M_{ij} := v_i^\dagger M v_j$. Note that we don't need $\mathcal{V}$ to be orthonormal to do all this.

Acknowledgements

These notes are mostly inspired from Scott Aaronson's undegraduate [Aar16b] and Ryan O'Donnell and John Wright's graduate, quantum computation notes [OW15].

1.1. Mathematical Preliminaries

We recall the notion of *tensor products*: Given $v \in \mathbb{C}^m, w \in \mathbb{C}^n$, $v \otimes w$ is a vector in $\mathbb{C}^{mn}$ given by:

$$v \otimes w := \begin{bmatrix} v_1 \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \\ \vdots \\ v_m \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \end{bmatrix}$$

One can think of the tensor product this way: If v, w represent probability distributions, then $v \otimes w$ represents the joint distribution of independent copies of v and w.
One can extend the notion of tensor products to general matrices too: Indeed, if $A \in \mathbb{C}^{a_1\times a_2}, B \in \mathbb{C}^{b_1\times b_2}$, then $A \otimes B \in \mathbb{C}^{a_1 b_1 \times a_2 b_2}$ is a matrix given by:

$$A \otimes B := \begin{bmatrix} a_{1,1}\begin{bmatrix} b_{1,1} & \cdots & b_{1,b_2} \\ \vdots & \ddots & \vdots \\ b_{b_1,1} & \cdots & b_{b_1,b_2} \end{bmatrix} & \cdots & a_{1,a_2}\begin{bmatrix} b_{1,1} & \cdots & b_{1,b_2} \\ \vdots & \ddots & \vdots \\ b_{b_1,1} & \cdots & b_{b_1,b_2} \end{bmatrix} \\ \vdots & \ddots & \vdots \\ a_{a_1,1}\begin{bmatrix} b_{1,1} & \cdots & b_{1,b_2} \\ \vdots & \ddots & \vdots \\ b_{b_1,1} & \cdots & b_{b_1,b_2} \end{bmatrix} & \cdots & a_{a_1,a_2}\begin{bmatrix} b_{1,1} & \cdots & b_{1,b_2} \\ \vdots & \ddots & \vdots \\ b_{b_1,1} & \cdots & b_{b_1,b_2} \end{bmatrix} \end{bmatrix}$$

Also note that not all vectors arise as tensor products (the probabilistic interpretation makes this very easy to see: Not all joint distributions are a product of their marginals): For example, the vector $\begin{bmatrix}0.5 & 0 & 0 & 0.5\end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^4$ is not the tensor product of any two vectors in $\mathbb{C}^2$.

1.2. Circuits

We want to develop a theory of computation for qubits (which we shall define shortly), and the first thing we need to define is the notion of a circuit.

Now, the Second Law of Thermodynamics says that irreversible boolean gates must dissipate energy: For example, take the AND gate. Given that the output of the AND gate is 0, there is no way for us to say if the inputs were $(0,0),(0,1)$ or $(1,0)$. Thus, the AND gate is not reversible, and any AND gate in nature must necessarily expend energy.
On the other hand, there is also the notion of universality of computation: Recall that NAND gates, along with *ancilla* bits, are universal for boolean computation. Indeed, given any boolean function $f : \{0,1\}^n \to \{0,1\}$, we can implement f by repeatedly applying NOT and AND gates to the input bits. Now, note that $\mathrm{AND}(x,y) = \mathrm{NOT}(\mathrm{NAND}(x,y))$. Finally, note that $\mathrm{NOT}(x) = \mathrm{NAND}(x,1)$. Note that we hardcoded one input of the NAND gate to be the bit 1; the constant bit 1 is called an *ancilla bit* in this computation.
Thus, one might wonder if there exist reversible gates that are also universal for boolean computation: The answer is yes, and the *Toffoli gate* is an example of such a gate, denoted as CCNOT.

Definition 1.1. The Toffoli gate $\mathrm{CCNOT} : \{0,1\}^3 \to \{0,1\}^3$ is defined as:

$$\mathrm{CCNOT}(x,y,z) := (x,y,\mathrm{AND}(x,y) \oplus z)$$

Remark. The formal definition of a reversible gate is as follows: Given the output of the gate, we should be able to uniquely determine the input of the gate. Consequently, a reversible gate as a function should be injective. Since we might as well restrict our co-domain to be the set of achievable outputs, it follows that reversible gates have to be bijective. In particular, they must necessarily have the same number of input and output bits.

Note that $\mathrm{CCNOT}(x,y,1) = (x,y,\mathrm{NAND}(x,y))$, and thus NAND gates can be simulated by the Toffoli gate; consequently, *Toffoli gates along with ancilla bits suffice for all boolean computation*.
Note that since CCNOT can simulate NOT, WLOG we can assume that all our ancilla bits are set to 0: We can transform any of them to 1 as required by applying a NOT gate.
We shall return to the issue of universality of computation later on.
Another particularly important gate in the context of quantum computing is the CNOT gate, which is a function from $\{0,1\}^2$ to $\{0,1\}^2$, which negates the second bit only if the first bit is 1, i.e. $\mathrm{CNOT}(0,x) = (0,x), \mathrm{CNOT}(1,x) = (1,1-x)$. Its matrix representation (note that functions $\{0,1\}^n \to \{0,1\}^n$ can be identified with matrices in $\{0,1\}^{2^n \times 2^n}$) is:

$$\mathrm{CNOT} := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

One interesting feature of the CNOT gate is that it can create correlations: For example, consider the vector $\begin{bmatrix} 0.5 & 0 & 0.5 & 0 \end{bmatrix}^\mathsf{T}$. It is the tensor product of $\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}^\mathsf{T}$ with $\begin{bmatrix} 1 & 0 \end{bmatrix}^\mathsf{T}$. Probabilistically, it represents that with probability $1/2$ our input is 00, and with probability $1/2$ it is 10. Now,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0 \\ 0.5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0 \\ 0 \\ 0.5 \end{bmatrix}$$

Note that the output is not expressible as a tensor product, i.e. it is *correlated*.

1.3. Qubits

A quantum state is a unit vector in $\mathbb{C}^N$ describing the state of a quantum system. The simplest interesting quantum system is the *qubit*, which has two levels, which we denote by $|0\rangle$ and $|1\rangle$. [1]

[1] a one-level quantum system would just be the state $|0\rangle$, which wouldn't be very interesting on its own. We could also talk about multi-level quantum systems, with states $|0\rangle, \ldots, |d\rangle$. However, we can simulate multi-level systems with qubits, so WLOG we only talk about qubits.

We now introduce the bra-ket notation: A vector $\psi \in \mathbb{C}^N$ is in the *ket-form*, and denoted as $|\psi\rangle$. The complex conjugate of ψ, is in the *bra-form*, and is denoted as $\langle\psi|$. For example, if $|\psi\rangle = \begin{bmatrix}\alpha\\ \beta\end{bmatrix}$, then $\langle\psi| = \begin{bmatrix}\alpha^* & \beta^*\end{bmatrix}$. Note that $\|\psi\|^2 = \langle\psi| \cdot |\psi\rangle$, which is also denoted as $\langle\psi|\psi\rangle$. Also note that $\langle x| \cdot |y\rangle =: \langle x|y\rangle = \langle y|x\rangle^*$.
We also define some very important quantum states:

Definition 1.2. We define:

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

$|\pm\rangle$ are also known as the Hadamard states.

Note that just as stochastic matrices act on probability distributions, the same way unitary matrices act on quantum states. Why unitary matrices? Quantum physics tells us that quantum transformations have to be linear, so only matrices can act on quantum states. But also note that quantum states have norm 1, which means that our matrix transformations will have to preserve norms, and consequently will have to be unitary. Recall that if $U \in \mathbb{C}^{N\times N}$ is unitary, then $U^\dagger U = UU^\dagger = I$, where $U^\dagger$ is the conjugate transpose of U.
Also note that unitary matrices are diagonalizable, which means that we can freely take square roots (or any positive real power) of a unitary matrix. For example, we can decompose the action of a unitary matrix U into the composition of actions of $\sqrt{U}$. This is quite natural, since in physics, unitary matrices denote the evolution of a system with time: Indeed, the Hamiltonian H of a quantum system is a Hermitian matrix, and if H is the Hamiltonian of a system, then the state of the system $|\psi\rangle$ after time t has elapsed is $e^{-iHt/\hbar}|\psi\rangle$, i.e. $U = e^{-iHt/\hbar}$. Clearly, we can write $U = e^{-iHt/\hbar} = e^{-iHt/2\hbar} \cdot e^{-iHt/2\hbar} = \sqrt{U} \cdot \sqrt{U}$, or $U = U^{\alpha_1} \cdot U^{\alpha_2} \cdots U^{\alpha_n}$, where $\alpha_1 + \cdots + \alpha_n = 1, \alpha_1, \ldots, \alpha_n \in [0, 1]$. This also illustrates why we need complex numbers for describing the evolution of a quantum state: Real unitary matrices, a.k.a orthogonal matrices, may not necessarily have real square roots.

1.4. Born's Rule

Suppose we have a quantum state $|\psi\rangle = \begin{bmatrix}\alpha_0\\ \alpha_1\end{bmatrix}$. The probability that we will see the qubit $|0\rangle$ when we *measure* ψ is $|\alpha_0|^2$. Similarly, the probability that we will see the qubit $|1\rangle$ when we measure ψ is $|\alpha_1|^2$. Since quantum states must have norm 1, we must have $|\alpha_0|^2 + |\alpha_1|^2 = 1$, which agrees with the laws of probability.
In general, suppose $V := \{|v_0\rangle, \ldots, |v_{n-1}\rangle\}$ is an orthonormal basis for our state space. [2] Then the probability we see the state $|v_i\rangle$ when we measure $|\psi\rangle$ against the basis V is $|\langle v_i|\psi\rangle|^2 = \langle\psi|v_i v_i^\dagger|\psi\rangle$. Note that measurements against orthonormal bases are known as *projective measurements*. We shall later see much more general forms of measurements.
What about partial measurement? Suppose our quantum state is $|\psi\rangle := \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$. Suppose we measure just the first qubit. The probability that we obtain $|0\rangle$ is $|\alpha_{00}|^2 + |\alpha_{01}|^2$. Now, suppose we do obtain $|0\rangle$ on measuring the first qubit. Then our quantum state $|\psi\rangle$ *collapses* into the state

$$\frac{\alpha_{00}}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}|00\rangle + \frac{\alpha_{01}}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}|01\rangle = |0\rangle \otimes \left(\frac{\alpha_{00}}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}|0\rangle + \frac{\alpha_{01}}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}|1\rangle\right)$$

If someone measures the second qubit now [3], the probability that they obtain $|0\rangle$ is $\frac{|\alpha_{00}|^2}{|\alpha_{00}|^2+|\alpha_{01}|^2}$.
We also make certain important remarks about the consequences of Born's rule:

1. Note that being certain in a measurement in the $\{|0\rangle, |1\rangle\}$ basis implies that we are maximally uncertain in our measurement in the $\{|+\rangle, |-\rangle\}$ basis, and vice versa.

[2] Why do we need V to be orthonormal? Note that to satisfy the laws of probability, we must have $\sum_{i=0}^{N-1} v_i v_i^\dagger = I \iff UU^\dagger = I$, where $U = \begin{bmatrix}|v_0\rangle & \cdots & |v_{N-1}\rangle\end{bmatrix}$, i.e. U is unitary, which means V is an orthonormal basis.
[3] with the knowledge that the first state was measured to be $|0\rangle$

2. Note that measurement is a very special type of quantum operation: Usual quantum operations are **reversible** (since unitary matrices are invertible), **deterministic** (given a state $|\psi\rangle$ and a unitary matrix U, applying U to $|\psi\rangle$ yields $U|\psi\rangle$: There is nothing probabilistic about the process of unitary evolution itself), and **continuous** (application of U can be decomposed into the application of $U^{1/n}$ n times, for any n). On the other hand, measurement itself is **irreversible** (once measured, we can't recover the superposition of a qubit back), **probabilistic** (the outcome of a measurement is not deterministic, unless we are measuring $|\psi\rangle$ against a basis containing $|\psi\rangle$), and **discontinuous** (the process of measurement is assumed to be instantaneous, i.e. the state of the system collapses instantaneously on measurement). Later on we shall see the *Superoperator formalism* which allows us to unite unitary operations and measurements into a single mathematical framework.

1.5. Quantum Interference

Note that a "mixing" of states exists even in classical computation: For example, a random bit can be written as $\frac{|0\rangle+|1\rangle}{2}$. However, classical randomness, and quantum superposition differ in two very crucial ways:

1. Even a random classical bit has a 'true' value, it is just that we are not completely sure what it is. A random bit has a definitive value in RAM, before we *measure* what that bit is. However, a quantum superposition is not like that: If our state is $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$, then this is truly a superposition until it is measured. It has no definitive value in the quantum computer before its measurement.

2. This leads us to our second point: Consider the transformation $U = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$. If we apply this to the qubit $|0\rangle$, we get the state $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$. However, note that $U = U^{-1}$, i.e. if we apply U again to $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$, then we get back $|0\rangle$! Note that this is not possible in a classical setting: If we apply a bijective function to a mixed (i.e. random) input, we can never obtain a deterministic output! The reason quantum computation allows us to 'unscramble' randomness is because the amplitudes of a state can be negative, and thus can cancel off positive amplitudes. This phenomenon is known as *interference*, and is the main cause of the power of quantum computation.

The above example also gives us this useful definition:

Definition 1.3 (Hadamard Gate)**.** The gate $H := \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$ is known as the *Hadamard gate*. Note that $H^2 = I$. Also note that $H|0\rangle = |+\rangle, H|1\rangle = |-\rangle$.

Remark. Note that H is the character table of $\mathbb{F}_2$. In some sense, Fourier analysis over finite groups is all the extra power that quantum computation has over classical computation.

1.6. Quantum Zeno Effect/Watched Pot Effect

The Quantum Zeno effect, first proposed by Alan Turing [Teu04], and later described rigorously in [DFG74], describes how we can arrest/induce change in a quantum system by measuring it continuously.
Suppose we have the qubit $|0\rangle$. Also, denote by V_θ the basis $\{v_\theta, v_\theta^\perp\}$, where $v_\theta := \cos(\theta)|0\rangle + \sin(\theta)|1\rangle, v_\theta^\perp := -\sin(\theta)|0\rangle + \cos(\theta)|1\rangle$. Suppose we measure $|0\rangle$ against V_ε. Then with probability $\cos^2(\varepsilon) \approx 1-\varepsilon^2$, the outcome of the measurement is v_ε. We can once again measure v_ε against $V_{2\varepsilon}$, and with probability $\approx 1-\varepsilon^2$, the outcome of the measurement will be $v_{2\varepsilon}$. Now, if we repeat this process $\frac{\pi}{2\varepsilon}$ times, and if we succeed every time, then we obtain the qubit $v_{\pi/2} = |1\rangle$! Also, the probability of failure is $\leq \mathcal{O}(\varepsilon^2)\cdot\frac{\pi}{2\varepsilon} = \mathcal{O}(\varepsilon)$.
Consequently, by letting $\varepsilon \searrow 0$, we can, with arbitrarily high probability, convert our $|0\rangle$ qubit into a $|1\rangle$ qubit by just measuring it repetitively against a drifting basis!

1.7. The Coin Problem

The quantum version of the coin problem was defined and solved by Aaronson and Drucker in [AD11].
Suppose we are given two coins, one which turns up 'Head' with probability $1/2$, and the other which turns up

'Head' with probability $\geq 1/2 + \varepsilon$. We have to identify the biased coin.
The classical way to do this is to toss both the coins $C \cdot \varepsilon^{-2}$ times, and keep track of the number of times each coin turns up 'Head'. With probability $\geq 1 - 2^{-\Omega(C)}$, the biased coin will have turned up 'Head' more number of times than the unbiased coin, which will give us the necessary distinction.
This algorithm requires $\mathcal{O}(\log 1/\varepsilon)$ bits to store the number of heads (if $C = \mathcal{O}(1)$). A theorem by Hellman and Cover [HC70] gives a corresponding lower bound of $\Omega(\log 1/\varepsilon)$ space for the coin problem.
Turns out that the above problem can be solved in $\mathcal{O}(1)$ space by a quantum algorithm [AD11]! We give a sketch of the algorithm here: Consider the rotation matrix $R_\theta := \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$. Initialize a qubit $|\psi\rangle := |0\rangle$. Pick any coin, and if it turns up heads apply R_ε to $|\psi\rangle$, and apply $R_{-\varepsilon}$ to $|\psi\rangle$ otherwise. Note that $R_{-\varepsilon} = R_\varepsilon^{-1}$. After repeating this process $\mathcal{O}(\varepsilon^{-2})$ times, we measure $|\psi\rangle$ (in the $\{|0\rangle, |1\rangle\}$ basis). If it turns up $|0\rangle$, w.h.p that coin was fair, otherwise not.
Thus, with qubits we can solve the coin problem in $\mathcal{O}(1)$ qubits, what in the classical setting necessarily takes $\Omega(\log 1/\varepsilon)$ bits!

1.8. Multi-Qubit Operations

Suppose we have a string of qubits $|x_1 x_2 \cdots x_n\rangle = |x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_n\rangle$. Say we want to apply the gate U_1 to (x_2, x_4, x_5), the gate U_2 to (x_1, x_3), and leave the rest of the qubits unchanged. Then the giant operator which acts on the entire string is $(U_1 \otimes U_2 \otimes I) \cdot \Pi$, where Π is a suitable permutation matrix which rearranges the string $|x_1 \cdots x_n\rangle$ to $|x_2 x_4 x_5\rangle \otimes |x_1 x_3\rangle \otimes |x_6 \cdots x_n\rangle$.

1.9. Entanglement

Apply the circuit $\mathrm{CNOT} \cdot (H \otimes I)$ to $|00\rangle$. Note that

$$H \otimes I = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

Multiplying the above with the CNOT gate yields the matrix

$$\mathrm{CNOT} \cdot (H \otimes I) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{bmatrix}$$

Thus, if we apply this to the state $|00\rangle = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^\mathsf{T}$, then we get the state $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix}^\mathsf{T}$, which corresponds to $|\rho\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.
The state $|\rho\rangle$ is also known as an Einstein-Podolsky-Rosen pair (EPR pair). The state $|\rho\rangle$ is entangled (i.e. can't be expressed as a tensor product).
Suppose the first qubit of the EPR pair is owned by Alice, and the second qubit is owned by Bob. When Alice measures her qubit, she will instantaneously know what Bob's measurement will yield, even though Alice and Bob could be physically separated by a large distance.
Einstein termed this *spooky action at a distance*. However, this is not surprising when one notes that correlated random variables are bound to behave this way: Indeed, suppose Alice and Bob are physically separated by a large distance, and they subscribe to the same newspaper. Alice knows that every morning the paper will either report that the markets have gone up, or down (but she doesn't know which of the two scenarios will happen). Then note that the moment Alice opens her paper, she also knows what Bob will read, and there is nothing spooky about this!
However, consider the following experiment: Suppose Alice measures her qubit in the $\{|+\rangle, |-\rangle\}$ basis. This is equivalent to first Hadamarding her qubit, and then measuring in the usual $\{|0\rangle, |1\rangle\}$ basis. That yields:

$$(H \otimes I)\left(\frac{|00\rangle + |11\rangle}{\sqrt{2}}\right) = \frac{|00\rangle + |01\rangle + |10\rangle - |11\rangle}{2}$$

Now, if Alice sees $|0\rangle$ on measuring her qubit, Bob's state collapses to $\frac{|00\rangle+|01\rangle}{\sqrt{2}} = |+\rangle$. Similarly, if Alice sees $|1\rangle$, then Bob's state collapses to $|-\rangle$.
Think for a moment about the implications of this: If Alice measures her qubit in the $\{|0\rangle, |1\rangle\}$ basis, then Bob's qubit collapses to $|0\rangle$ or $|1\rangle$. But if Alice measures her qubit in the $\{|+\rangle, |-\rangle\}$ basis, then Bob's qubit collapses to $|+\rangle$ or $|-\rangle$! Then the question arises: How does Bob's qubit know which basis Alice measured her qubit in, so that it knows how to collapse to the "correct" set of states? We use the formalism of *mixed states* to answer this.
Before we move on to mixed states, we mention a last weird consequence of entanglement: Suppose we have an EPR pair $|\psi\rangle := \frac{|00\rangle+|11\rangle}{\sqrt{2}}$. We place another qubit $|0\rangle$ with $|\psi\rangle$, and thus the state of the system $|\psi\rangle \otimes |0\rangle$. We now swap the last two qubits to obtain the state $\frac{|000\rangle+|101\rangle}{\sqrt{2}}$. Note that the third qubit got entangled with the first qubit without these qubits ever interacting with each other, i.e. entanglement is transferable!

1.10. Mixed States and Density Matrices

Given a set of quantum states $S := \{|\psi_i\rangle\}_{i\in\mathcal{I}}$, we define a *mixed state* as a probability distribution over these quantum states, i.e. a mixed state is the tuple $\{(p_i, |\psi_i\rangle)\}_{i\in\mathcal{I}}$, with $\sum_{i\in\mathcal{I}} p_i = 1$, $p_i \in [0, 1]$. The states $|\psi_i\rangle$ are called *pure states* in this context. A pure state can also be considered as a mixed state, where the corresponding probability distribution has a support of size 1.
The *density matrix* corresponding to a mixed state is defined to be:

$$\rho := \sum_{i\in\mathcal{I}} p_i |\psi_i\rangle\langle\psi_i|$$

1.10.1. Properties of The Density Matrix

We quickly list down some properties of the Density Matrix:

1. **Hermitian**: It is easy to see that $\rho^\dagger = \rho$.
2. **Trace** 1: Note that $\mathrm{tr}(\rho) = \sum_{i\in\mathcal{I}} p_i \sum_{j=1}^N \psi_{ij}\psi_{ij}^* = \sum_{i\in\mathcal{I}} p_i \sum_{j=1}^N |\psi_{ij}|^2 = \sum_{i\in\mathcal{I}} p_i \|\psi_i\|^2 = \sum_{i\in\mathcal{I}} p_i = 1$.
3. **Positive Semi-Definite (PSD)**: For any $|x\rangle$, $x^\dagger\rho x = \langle x|\rho|x\rangle = \sum p_i\langle x||\psi_i\rangle\langle\psi_i||x\rangle = \sum p_i|\langle\psi_i|x\rangle|^2 \geq 0$.

Conversely, let ϱ be any Hermitian PSD matrix with trace 1. Since ϱ is Hermitian, the spectral theorem applies, and we can write $\varrho = \sum_{j=1}^N \lambda_j |v_j\rangle\langle v_j|$, where $\{v_j\}$ are eigenvectors of ϱ, and $\{\lambda_j\}$ are the corresponding eigenvalues. Since ϱ is Hermitian, $\lambda_j \in \mathbb{R}$, and since ϱ is PSD, $\lambda_j \geq 0$. Finally, $\sum_{j=1}^N \lambda_j = \mathrm{tr}(\varrho) = 1$, and thus $\{\lambda_j\}$ is a collection of non-negative real numbers which sum to 1. Consequently, $\{\lambda_j\}$ can be viewed as a probability distribution, and thus ϱ becomes the density matrix of the mixed state $\{(\lambda_j, |v_j\rangle)\}_{j\in[N]}$. Summarizing the above discussion, we get the following theorem:

Theorem 1.1. A square matrix ρ is a density matrix if and only if it is Hermitian, PSD, and has trace 1.

We list some further properties of the density matrix:

1. **Rank of the Density Matrix**: Suppose ρ is the density matrix of some mixed state. Let $r = \mathrm{rank}(\rho)$. Then note that ρ is also the density matrix of a mixed state with r pure components. In particular, any density matrix arises as the density matrix of a mixed state with $\leq N$ pure components. Furthermore, the rank of a density matrix tells us the minimum number of pure states we have to mix to achieve the given density matrix. Also note that a density matrix represents a pure state if and only if it has rank 1.
2. **Maximally Mixed State**: Suppose $\{|\psi_1\rangle, \ldots, |\psi_N\rangle\}$ are an **orthonormal basis** for our state space. Consider the mixed state $\left\{\left(\frac{1}{N}, |\psi_i\rangle\right)\right\}$. Then the density matrix of this state is $\frac{1}{N}\sum_{i=1}^N |\psi_i\rangle\langle\psi_i|$. Now, recall from linear algebra that if $\{|\psi_1\rangle, \ldots, |\psi_N\rangle\}$ is an orthonormal basis, then $\sum_{i=1}^N |\psi_i\rangle\langle\psi_i| = I$. Consequently, the density matrix for the uniform distribution on any orthonormal basis is I/N. If the density matrix of a N-dimensional Hilbert space is I/N, then it is called a *maximally mixed state*.

3. **Effect of Unitary Transformations**: Suppose we apply a unitary transformation to all the pure components of a mixed state with density matrix ρ. Then the new density matrix becomes $\sum_{i\in\mathcal{I}} p_i(U|\psi_i\rangle)(U|\psi_i\rangle)^\dagger = U\rho U^\dagger$.

4. **Pure vs. Mixed States**: Let ρ be a density matrix. Diagonalize $\rho = UDU^\dagger$. Then $\rho^2 = UD^2U^\dagger$, and $\mathrm{tr}(\rho^2) = \mathrm{tr}(D^2) = \sum \lambda_i^2$, where $\{\lambda_i\}$ are the eigenvalues of ρ. Now, $\sum \lambda_i^2 \leq \sum \lambda_i = 1$, with equality occuring only if $\lambda_i = 0, 1$ for all i. But that happens only when ρ is pure. Thus $\mathrm{tr}(\rho^2) \leq 1$ for all density matrices, with equality occuring if and only if ρ is pure. Thus, this can be used as a criteria to judge if a state is pure/mixed from its density matrix.

5. **Measurements**: Suppose we have a mixed state $\tau := \{(p_i, |\psi_i\rangle)\}_{i\in\mathcal{I}}$, with density matrix $\rho = \sum_{i\in\mathcal{I}} p_i|\psi_i\rangle\langle\psi_i|$. Also suppose we have an orthonormal basis $V := \{|v_j\rangle\}_{j\in[N]}$. What happens if we measure τ against V?

$$\Pr(\text{We observe } v_j) = \sum_{i\in\mathcal{I}} p_i|\langle v_j|\psi_i\rangle|^2 = \sum_{i\in\mathcal{I}} p_i\langle v_j|\psi_i\rangle\langle\psi_i|v_j\rangle = \left\langle v_j \middle| \sum_{i\in\mathcal{I}} p_i|\psi_i\rangle\langle\psi_i| \middle| v_j \right\rangle = \langle v_j|\rho|v_j\rangle$$

   Consequently, *the results of any measurement on $|\tau\rangle$ are completely captured by its density matrix*. In particular, if two mixed states have the same density matrix, then **no (projective) measurement can distinguish them**. In other words, the density matrix of a quantum system provides a complete description of it.

6. **Off-Diagonal Entries**: Note that $|0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. Also note that $|+\rangle\langle +| = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$. The off-diagonal entries of the density matrix of $|0\rangle$ are 0, while they are non-zero for $|+\rangle\langle +|$. This is because off-diagonal entries represent the "interference" in the system. Since the qubits $|0\rangle$ and $|1\rangle$ are maximally superimposed in $|+\rangle$, there is a lot of interference between them, which shows up in the off-diagonal entries of the density matrix.
   However, note that the off-diagonal entries of ρ are basis dependent. For example, ρ is a diagonal matrix in its own eigenbasis, i.e. since ρ is Hermitian, we can diagonalize ρ as $UDU^\dagger$, and in the basis given by the columns of U, ρ is a diagonal matrix.

7. **Diagonal of the Density Matrix**: Suppose ρ, as a matrix, is expressed in the orthonormal basis V. Then note that the diagonal entries of ρ represent the probabilities of obtaining various basis vectors of V when measuring ρ (against V). More precisely, $\rho_{ii} = \Pr(\text{We obtain } |i\rangle \text{ on measuring } \rho)$. Thus, if $\rho \in \mathbb{C}^{N\times N}$, then $\mathrm{diag}(\rho) \in [0,1]^N \hookrightarrow \mathbb{C}^N$ is a probability distribution on V.

1.10.2. The Reduced Density Matrix

This section is taken from the Wikipedia page on partial traces [Wik24].
Let $\mathcal{H}$ be the Hilbert space our system lies in. If we are dealing with a system of n qubits, then $\mathcal{H}$ is just $\mathbb{C}^{2^n}$. Now suppose Alice owns a qubits among those n, and Bob owns the other $b := n - a$ qubits. Define $\mathcal{H}_A \cong \mathbb{C}^{2^a}$ to be the subsystem of Alice, and $\mathcal{H}_B \cong \mathbb{C}^{2^b}$ to be the subsystem of Bob. Clearly $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$.
For any Hilbert space $\mathcal{H}$, define $L(\mathcal{H})$ to be the space of linear operators over $\mathcal{H}$. Note that if $\mathcal{H} \cong \mathbb{C}^N$, then $L(\mathcal{H}) \cong \mathbb{C}^{N\times N}$. Then we define the *partial trace* operator $\mathrm{tr}_B : L(\mathcal{H}_A \otimes \mathcal{H}_B) \to L(\mathcal{H}_A)$ as the unique linear map such that $\mathrm{tr}_B(R \otimes S) = \mathrm{tr}(S)R$ for all $R \in L(\mathcal{H}_A), S \in L(\mathcal{H}_B)$. To show the uniqueness of tr_B, it suffices to produce a basis for $L(\mathcal{H}_A \otimes \mathcal{H}_B)$, and show that tr_B specifies a mapping for all the basis elements.
To that extent, let $\{|i\rangle_A\}$ be *any* basis for $\mathcal{H}_A$, and let $\{|j\rangle_B\}$ be any basis for $\mathcal{H}_B$. Then $\{|i\rangle_A\langle i'|_A\}$ is a basis for $L(\mathcal{H}_A)$, $\{|j\rangle_B\langle j'|_B\}$ is a basis for $L(\mathcal{H}_B)$, and thus $\mathcal{B} := \{|i\rangle_A\langle i'|_A \otimes |j\rangle_B\langle j'|_B\}$ is a basis for $L(\mathcal{H})$. But note that the rule '$\mathrm{tr}_B(R \otimes S) = \mathrm{tr}(S)R$' specifies the value of tr_B for all elements of $L(\mathcal{H}_A) \otimes L(\mathcal{H}_B)$, and thus for all elements of the basis $\mathcal{B}$ for $L(\mathcal{H})$, as desired.
Note that we can define tr_A similarly as above. Finally, the name "partial trace" should also be clear: Indeed, if $b = 0$, then $\mathrm{tr}_A = \mathrm{tr}$.
We can now finally define the notion of reduced density matrix.

Definition 1.4. Suppose ρ is the density matrix of a quantum state (possibly mixed) in the Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Then the **Reduced Density Matrix** of Alice is defined to be $\rho_A := \mathrm{tr}_B(\rho)$, and the Reduced Density Matrix of Bob is defined to be $\rho_B := \mathrm{tr}_A(\rho)$.
The process of taking partial traces is also known as *tracing out*, as in trace "out" the *other* subsystem.

How do we explicitly calculate the Reduced Density Matrix? Let $\{|i\rangle_A\}$ be any basis for $\mathcal{H}_A$, and let $\{|j\rangle_B\}$ be any basis for $\mathcal{H}_B$. Note that $\{|i\rangle_A\}, \{|j\rangle_B\}$ need not be orthonormal. Let $|\psi\rangle$ be a pure state. Then, for some coefficients α_{ij}, we have:

$$|\psi\rangle = \sum_{i,j} \alpha_{ij}|i\rangle_A \otimes |j\rangle_B = \sum_i |i\rangle_A \otimes \left(\sum_j \alpha_{ij}|j\rangle_B\right) =: \sum_i |i\rangle_A \otimes |v_i\rangle_B$$

$$\implies |\psi\rangle\langle\psi| = \sum_{i,i'} (|i\rangle_A \otimes |v_i\rangle_B) \cdot (\langle i'|_A \otimes \langle v_{i'}|_B) = \sum_{i,i'} |i\rangle_A\langle i'|_A \otimes |v_i\rangle_B\langle v_{i'}|_B$$

$$\implies \mathrm{tr}_A(|\psi\rangle\langle\psi|) = \sum_{i,i'} \mathrm{tr}_A\left(|i\rangle_A\langle i'|_A \otimes |v_i\rangle_B\langle v_{i'}|_B\right) = \sum_{i,i'} \mathrm{tr}\left(|i\rangle_A\langle i'|_A\right)|v_i\rangle_B\langle v_{i'}|_B = \sum_{i,i'} \delta_{i,i'}|v_i\rangle_B\langle v_{i'}|_B$$

$$\implies \mathrm{tr}_A(|\psi\rangle\langle\psi|) = \sum_i |v_i\rangle_B\langle v_i|_B$$

Thus, if $\rho = |\psi\rangle\langle\psi|$ is the density matrix of a pure state, then $\rho_B = \sum_i |v_i\rangle_B\langle v_i|_B$. In case ρ is the density matrix of a mixed state, we can calculate the reduced density matrices for each of the pure components as above, and then take a weighted linear combination of those matrices to obtain the Reduced Density Matrix for Bob.
Note that

$$\sum_i |v_i\rangle_B\langle v_i|_B = \sum_i \sum_{j,k} \alpha_{ij}\alpha^*_{ik}|j\rangle_B\langle k|_B$$

Thus, if we express ρ_B in the $\{|j\rangle_B\}$ basis, then $(\rho_B)_{jk} = \sum_i \alpha_{ij}\alpha^*_{ik}$. Now, denote by α the $2^a \times 2^b$ matrix whose $(i,j)^{\text{th}}$ entry is α_{ij}. Then $(\rho_B)_{jk} = \sum_i \alpha_{ij}\alpha^*_{ik} = \sum_i \alpha_{ij}(\alpha^\dagger)_{ki} = (\alpha^\dagger\alpha)_{kj}$. Thus $\rho_B = (\alpha^\dagger\alpha)^\mathsf{T}$.
Now, suppose Alice performs a unitary transformation U on her qubits. Then our new state is

$$|\psi'\rangle = \sum_{i,j} \alpha_{ij}(U|i\rangle_A) \otimes |j\rangle_B$$

Now, $U|i\rangle_A = \sum_{i'} U_{i'i}|i'\rangle_A$, where we express U in the $\{|i\rangle_A\}$ basis. Thus

$$|\psi'\rangle = \sum_{i',j} \left(\sum_i U_{i'i}\alpha_{ij}\right)|i'\rangle_A \otimes |j\rangle_B = \sum_{i',j} (U\alpha)_{i'j}|i'\rangle_A \otimes |j\rangle_B = \sum_{i,j} (U\alpha)_{ij}|i\rangle_A \otimes |j\rangle_B$$

Thus the new density matrix $((U\alpha)^\dagger U\alpha)^\mathsf{T} = (\alpha^\dagger U^\dagger U\alpha)^\mathsf{T} = (\alpha^\dagger\alpha)^\mathsf{T} = \rho_B$.
Now, assume $\{|i\rangle_A\}, \{|i\rangle_A \otimes |v_i\rangle_B\}$ are orthonormal bases, and suppose Alice performs a measurement on her qubits in the $\{|i\rangle_A\}$ basis. What happens to ρ_B when she does that? Note that Alice obtains the state $|i\rangle_A$ with probability $p_i := \sum_j |\alpha_{ij}|^2$. From Bob's point of view, Bob doesn't know what Alice's measurement yields. Thus, according to Bob, Alice now has the mixed state

$$\left\{\left(p_i, \frac{|i\rangle_A \otimes |v_i\rangle_B}{\sqrt{p_i}}\right)\right\}$$

Thus, the Reduced Density Matrix which Bob sees is

$$\sum_i p_i \frac{|v_i\rangle_B}{\sqrt{p_i}} \cdot \frac{\langle v_i|_B}{\sqrt{p_i}} = \rho_B$$

Thus, Bob's Reduced Density Matrix doesn't change even if Alice performs a measurement on her qubits. Consequently,

Theorem 1.2 (No-Communication Theorem)**.** Let $|\psi\rangle$ be any state (possibly mixed). Suppose Alice has some qubits of $|\psi\rangle$, and suppose Bob has the rest. Then Bob's reduced density matrix doesn't change, regardless of whether Alice performs any measurement (on any orthonormal basis), or applies any unitary transformation to her qubits.

Remark. Although we proved the above theorem for pure states only, the mixed state analog goes through as it is. Also note that the above theorem permits measurement on any basis, not just the canonical one: However, measurement on an arbitrary orthonormal basis can be simulated as first applying a unitary transformation on the qubits, and then measuring in the usual basis. Since both of these operations preserve the reduced density matrix, so does their composition.

Why is the above theorem called the No-Communication theorem? That is because regardless of what Alice does with her qubits, Bob can't know that, since all the information Bob has is encapsulated in his Reduced Density Matrix, and that doesn't change with Alice's actions. Thus the formalism of mixed states, along with the notion of Reduced Density Matrices, provides a resolution of the EPR paradox: Faster-than-light communication doesn't happen, since no information is conveyed to Bob through Alice's actions.

1.10.3. POVM and Superoperator Formalism

This section is taken from [Aar16a].

Let $\{v_1, \dots, v_N\}$ be a basis of our Hilbert space $\mathcal{H} \cong \mathbb{C}^N$. If $|\psi\rangle \in \mathbb{C}^N$ is a quantum state, then the probability we obtain v_i on measuring $|\psi\rangle$ is $\langle\psi|v_i v_i^\dagger|\psi\rangle$. Now, note that there can be more general forms of measurements: Indeed, suppose we have the state $|\psi\rangle$. We add the qubit $|0\rangle$ to the system, making the state $|\psi\rangle \otimes |0\rangle$. Now, when we measure this state, we will get a variety of outcomes, but they won't be basis of $\mathbb{C}^{2N}$.

We thus introduce the so-called *Positive Operator-Valued Measurements*: Indeed, let $\{E_1, \dots, E_k\}$ [4] be a collection of $N \times N$ Hermitian PSD operators, with the property that $\sum E_i = I$. Then the probability we obtain the *observable* E_i on measuring the density matrix ρ is given by $\operatorname{tr}(E_i\rho)$. This also explains the quantum mechanics point-of-view wherein Hermitian PSD operators are treated as "observables".

To view projective measurements as POVMs, set $E_i = v_i v_i^\dagger$. Then for a pure state $|\psi\rangle$, $\operatorname{tr}(v_i v_i^\dagger \rho) = \operatorname{tr}(v_i v_i^\dagger \psi\psi^\dagger) = \operatorname{tr}(\psi^\dagger v_i v_i^\dagger \psi) = \langle\psi|v_i v_i^\dagger|\psi\rangle$, as desired. Conversely, one can view all POVMs as projective measurements on states augmented with ancilla qubits.

There is an even more general formalism known as the *Superoperator formalism*: Given a density matrix $\rho \in \mathbb{C}^{N\times N}$, and an set of matrices $A_1, \dots, A_\ell \in \mathbb{C}^{N\times M}$ satisfying $\sum A_i A_i^\dagger = I$, the superoperator maps

$$\rho \mapsto \sum A_i^\dagger \rho A_i$$

It is easy to verify that $\sum A_i^\dagger \rho A_i$ is a valid $M \times M$ density matrix.

Now, in projective measurements, we measure our system to obtain the states in the orthonormal basis $\{|v_i\rangle\}_{i\in[N]} \subseteq \mathbb{C}^N$, or alternatively the "observables" $|v_i\rangle\langle v_i|$. The observable $|v_i\rangle\langle v_i|$, corresponding to the state $|v_i\rangle$, is observed with probability $\langle v_i|\rho|v_i\rangle = \operatorname{tr}(\langle v_i|\rho|v_i\rangle) = \operatorname{tr}(|v_i\rangle\langle v_i|\rho)$. In the POVM formalism, we dispense away with the restriction that $\{|v_i\rangle\}$ need to be orthonormal. We can now measure against arbitrary sets $\{|w_i\rangle\}_{i\in[k]} \subseteq \mathbb{C}^N$, provided that we have $\sum_{i=1}^k E_i = I$, where $E_i = |w_i\rangle\langle w_i|$ is the observable corresponding to the state $|w_i\rangle$. [5] The probability of observing the observable E_i, corresponding to the state $|w_i\rangle$, on measurement is $\operatorname{tr}(|w_i\rangle\langle w_i|\rho) = \operatorname{tr}(E_i\rho)$, as desired.

The superoperator formalism goes one step further: What if our "states" are now *rectangular matrices*, instead of vectors, like $|v_i\rangle$ or $|w_i\rangle$? Indeed, let the set of states be $\{A_1, \dots, A_\ell\} \subseteq \mathbb{C}^{N\times M}$, satisfying the usual requirement that $\sum_{i=1}^\ell A_i A_i^\dagger = I$. Then the superoperator formalism says that we first transform our density matrix itself into a new density matrix, as dictated by the mapping $\rho \mapsto \sum A_i^\dagger \rho A_i$. Then the probability of obtaining the "state" A_i, or observing the observable $A_i A_i^\dagger$, on measuring ρ is $\operatorname{tr}(A_i^\dagger \rho A_i) = \operatorname{tr}(A_i A_i^\dagger \rho)$, i.e. the traces of the various constituents of the new density matrix actually encode probabilities of observing the corresponding observables. Furthermore, if we observe $A_i A_i^\dagger$, then the density matrix of our system collapses from ρ to $A_i^\dagger \rho A_i / \operatorname{tr}(A_i^\dagger \rho A_i)$.

[4] Note that k may be smaller than, equal to, or greater than N. In particular, if $k > N$, then there are more than N outcomes on measuring ρ

[5] Note that any Hermitian PSD matrix E can be factorized as $|w\rangle\langle w|$

Note that if we apply a unitary transformation U to ρ, the new density matrix we get is $U\rho U^\dagger$. Thus, the superoperator formalism encompasses within itself unitary transformations, measurements, and ancilla qubits. Furthermore, quantum physics shows that applying a superoperator to a density matrix is the only allowable quantum transformation, thus establishing this formalism to be the most general. Also note that while the appearance of the density matrix in Item 5 was a happy consequence of Born's rule, in the Superoperator formalism all our manipulations are w.r.t. the density matrix only, i.e. we completely ignore the "internal" specifics of whether our state is pure or not. Thus, in the superoperator formalism, the density matrix provides a complete description of our system, *by fiat*.

1.10.4. Purifications and the Schrödinger-HJW Theorem

Let $\rho \in \mathbb{C}^{N\times N}$ be a density matrix. Let $\rho = \sum_{i=1}^N \lambda_i |v_i\rangle\langle v_i| = \sum_{i=1}^N |w_i\rangle\langle w_i|$ be the spectral decomposition of ρ, where $|w_i\rangle := \sqrt{\lambda_i}|v_i\rangle$. Note that $\{|w_j\rangle\}$ forms a basis of $\mathbb{C}^N$. Let $\{|u_i\rangle_A\}_{i\in[N]}$ be an arbitrary basis of $\mathbb{C}^N$. Then note that $\rho = \mathrm{tr}_A(|\psi\rangle\langle\psi|)$, where $\psi := \sum \alpha_{ij}|u_i\rangle_A \otimes |w_j\rangle \in \mathbb{C}^{N^2}$, where the coefficients α_{ij} are chosen such that $\|\psi\|_2 = 1$ [6]. Consequently, we have the following definition and theorem:

Theorem 1.3 (Purification)**.** Let ρ be a density matrix. A pure state $|\psi\rangle$ such that $\rho = \mathrm{tr}_A(|\psi\rangle\langle\psi|)$ is known as a purification of ρ.

Theorem 1.4 (Schrödinger-HJW Theorem)**.** Let $\rho \in \mathbb{C}^{N\times N}$ be (the density matrix of) a mixed state. Then there exists a purification $\psi \in \mathbb{C}^M$ of $M \leq N^2$ dimensions such that $\rho = \mathrm{tr}_A(|\psi\rangle\langle\psi|)$.

1.11. No-Cloning Theorem

Suppose there was some circuit for *cloning* qubits, i.e. given as input a qubit (and some ancilla bits), the circuit would output two unentangled copies of the same qubit.
If such a circuit existed, then in the EPR experiment, Bob could prepare a large number of copies of his qubit, and measure them, and consequently know if Alice had measured her qubit in $\{|0\rangle, |1\rangle\}$ basis or the $\{|+\rangle, |-\rangle\}$ basis, thus enabling faster-than-light communication.
Thus, if the laws of physics hold, then such a cloning circuit shouldn't exist. Let us also see a mathematical proof of the fact.

Theorem 1.5 (No-Cloning Theorem)**.** There doesn't exist a circuit which clones a given qubit.

Proof. Suppose there did exist such a circuit. Writing the circuit as a unitary transform, we get that $U(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle$, where $|0\rangle$ is an ancilla qubit.
Thus, if $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, then

$$U \cdot \begin{bmatrix} \alpha \\ 0 \\ \beta \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha^2 \\ \alpha\beta \\ \alpha\beta \\ \beta^2 \end{bmatrix}$$

for all $\alpha, \beta \in \mathbb{C}$ such that $|\alpha|^2 + |\beta|^2 = 1$. But the above transformation is not linear, and we thus have a contradiction. ■

[6] note that the choice of these coefficients is not unique

1.12. Deferred Measurement Principle

Let $\mathcal{C}$ be a quantum circuit. Then the Deferred Measurement Principle says that WLOG we can assume that all measurements in $\mathcal{C}$ are carried out in the end (after all the unitary transformations have been applied), at the cost of adding a few extra ancillary qubits in the beginning.
Indeed, suppose we're measuring the qubits $\{i_1, \ldots, i_r\}$ in the middle of the circuit. Let $\mathcal{H}_m$ be the Hilbert space corresponding to these qubits, and let $\mathcal{H}$ be the Hilbert space of the whole system. Then $\mathcal{H} = \mathcal{H}_m \otimes \mathcal{H}_u$, where $\mathcal{H}_u$ is the Hilbert space corresponding to the unmeasured qubits.
Now, suppose the state of the system before the measurement is $|\psi_1\rangle = \sum_{i,j} \alpha_{ij} |i\rangle_m \otimes |j\rangle_u$, where $\{|i\rangle_m\}$ is the basis of $\mathcal{H}_m$ against which measurements are carried out, and $\{|j\rangle_u\}$ is any basis of $\mathcal{H}_u$. Also, let $|\psi_2\rangle = \sum_{i,j} \alpha'_{ij} |i\rangle_m \otimes |j\rangle_u$ is the state of the system after the measurement and the rest of the circuit is executed.
Now, consider the alternative circuit, where we introduce the ancillary qubits $|0\rangle^{\otimes r}$ in the beginning, whose corresponding Hilbert space is $\mathcal{H}_a$. Before and after the measurement against $\mathcal{H}_m$ is carried out, the unitary transform acting on these qubits is Id. Now, instead of measuring $|\psi_1\rangle$ against $\mathcal{H}_m$, we apply a CNOT type operation, which acts as:

$$|i\rangle_m \otimes |\psi'\rangle_u \otimes |0\rangle^{\otimes r} \mapsto |i\rangle_m \otimes |\psi'\rangle_u \otimes |i\rangle_m$$

After that, we execute the rest of the circuit as it is. In the end, we first measure the ancillary qubits. If the ancillary qubits are measured to be $|i\rangle_m$, then the remaining qubits collapse to $|\psi_2\rangle$, exactly as they would have in the original circuit! Thus, by entangling the ancillary qubits with the measured qubits, we can replace the measurement with a unitary operation.

1.13. Discarding a Qubit is equivalent to Measuring it

Suppose we have a state $|\psi\rangle = \alpha|0\rangle \otimes |\psi_0\rangle + \beta|1\rangle \otimes |\psi_1\rangle$. Suppose we measure the first qubit. Then the remaining system collapses into a mixed state: With probability $|\alpha|^2$ it collapses to $|\psi_0\rangle$, and with probability $|\beta|^2$ it collapses to $|\psi_1\rangle$. Thus the reduced density matrix of the remaining qubits is $|\alpha|^2|\psi_0\rangle\langle\psi_0| + |\beta|^2|\psi_1\rangle\langle\psi_1|$.
On the other hand the density matrix of $|\psi\rangle$ is $\rho = |\psi\rangle\langle\psi| = |\alpha|^2|0\rangle\langle 0| \otimes |\psi_0\rangle\langle\psi_0| + \alpha\beta^*|0\rangle\langle 1| \otimes |\psi_0\rangle\langle\psi_1| + \alpha^*\beta|1\rangle\langle 0| \otimes |\psi_1\rangle\langle\psi_0| + |\beta|^2|1\rangle\langle 1| \otimes |\psi_1\rangle\langle\psi_1|$. Tracing out the first qubit yields the reduced density matrix to be $|\alpha|^2|\psi_0\rangle\langle\psi_0| + |\beta|^2|\psi_1\rangle\langle\psi_1|$. Consequently, there is no difference between discarding a qubit and measuring it.

1.14. Uncomputation

Ancillary qubits are not as harmless as they seem: Indeed, if we discard them at the end of the computation, that is equivalent to measuring them. However, ancillary qubits may be entangled with the qubits in which we're interested in, and thus discarding the ancillary qubits may cause our desired qubits to collapse into a certain subset of possibilities, when we might have instead been interested in dealing with the whole superposition itself.
The way we deal with this is using a technique called uncomputation: Suppose a unitary transforms $|x\rangle \otimes |0\rangle^a$ to $|\phi\rangle \otimes |\,\mathrm{garbage}(x)\rangle \otimes |f(x)\rangle$, i.e. $U|x\rangle \otimes |0\rangle^a = |\phi\rangle \otimes |\,\mathrm{garbage}(x)\rangle \otimes |f(x)\rangle$. We're interested in getting rid of the garbage qubits without collapsing the remaining qubits.
The trick to doing that is to first add some ancillary qubits to $|\phi\rangle \otimes |\,\mathrm{garbage}(x)\rangle \otimes |f(x)\rangle$ to make it $|\phi\rangle \otimes |\,\mathrm{garbage}(x)\rangle \otimes |f(x)\rangle \otimes |0\rangle^r$, where r is the number of qubits in $f(x)$. Then we apply a CNOT type operation to transform $|\phi\rangle \otimes |\,\mathrm{garbage}(x)\rangle \otimes |f(x)\rangle \otimes |0\rangle^r \mapsto |\phi\rangle \otimes |\,\mathrm{garbage}(x)\rangle \otimes |f(x)\rangle \otimes |f(x)\rangle$. Finally, apply $U^{-1} \otimes \mathrm{Id}$ to the whole state, to yield

$$(U^{-1} \otimes \mathrm{Id}) \cdot |\phi\rangle \otimes |\,\mathrm{garbage}(x)\rangle \otimes |f(x)\rangle \otimes |f(x)\rangle = |x\rangle \otimes |0\rangle^a \otimes |f(x)\rangle$$

At this point, even if we discard $|0\rangle^a$, we're still left with $|x\rangle \otimes |f(x)\rangle$, as desired.

Remark. Note that in uncomputation, we assume that the function f is classical. Indeed, if f were a quantum function, then we wouldn't be able to clone it as $|f(x)\rangle \otimes |0\rangle^r \mapsto |f(x)\rangle \otimes |f(x)\rangle$ (due to the No-Cloning theorem). Indeed, getting rid of garbage for quantum f is a very subtle, tricky, and not-yet-fully resolved issue.

1.15. Miscellaneous Points about Quantum Computing

Note that a quantum computer with n qubits can be simulated by a classical computer with $2^n \,\mathrm{poly}(n)$ bits. Thus any quantum algorithm can at most give us an exponential speedup over classical computers. In particular, if some function is not computable in the classical sense, then it remains uncomputable in the quantum sense.

Note that without entanglement, we can simulate quantum computation classically with polynomial overhead: Indeed, without entanglement, all states we'd be dealing with would be product states of the form $(\alpha_1|0\rangle + \beta_1|1\rangle) \otimes \cdots \otimes (\alpha_n|0\rangle + \beta_n|1\rangle)$, which can be simulated with $2n \cdot \text{poly}(n) = \text{poly}(n)$ classical bits. Thus entanglement is key to obtaining quantum speedups.

Any n-qubit unitary transformation can be described as a circuit of 1 and 2-qubit gates. However, this procedure may take $\sim 4^n$ gates to build. Since we're interested in poly-time computation, we will be interested in unitaries which can be built with poly-number of gates. Indeed, the set of all n-qubit unitaries forms a $2^{\mathcal{O}(n)}$-dimensional manifold. On the other hand, a quantum circuit with T gates, each gate taking at most 2 qubits in, can be specified using $\mathcal{O}(T)$ continuous parameters. Thus, in the Haar measure sense, $1 - o_n(1)$ fraction of all n-qubit unitaries are not constructible using poly-sized circuits. In fact, even if we're content with just approximating [7] n-qubit unitaries, even then $1 - o(1)$ fraction of all n-qubit unitaries remain out of reach of poly-sized circuits.

We call a set $\mathcal{S}$ of gates *universal* if any n-bit unitary can be approximated to arbitrary precision by composing gates from $\mathcal{S}$, i.e. the set of unitaries constructible from $\mathcal{S}$ is dense [8] in the set of all unitaries.

As we know, $\{\text{Toffoli}\}$ is universal for classical boolean computation. A result due to Shi says that $\{\text{Toffoli}, \text{Hadamard}, S\}$, where $S := \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$, is universal for quantum computation. Note that S is needed for universality since $\{\text{Toffoli}, \text{Hadamard}\}$ can only construct real-valued unitaries.

The construction/approximation of unitaries from universal gates sets can also be made efficient through the Solovay-Kitaev theorem: The **Solovay-Kitaev theorem** says that if $\mathcal{S}$ is a universal gate set **closed under inverses**, i.e. $G \in \mathcal{S} \implies G^{-1} \in \mathcal{S}$, then any n-qubit unitary can be approximated within an operator norm of ε using only $2^{\mathcal{O}(n)} \log^{\mathcal{O}(1)}\left(\frac{1}{\varepsilon}\right)$ gates from $\mathcal{S}$. Thus, if n is held constant, then the number of gates required doesn't scale too badly with ε. Moreover, the circuit achieving the Solovay-Kitaev bound can also be found reasonably quickly.

1.15.1. Unitary Synthesis Problem

Note that Shannon's theorem says that most (i.e. $1 - o_n(1)$) n-bit Boolean functions take $\Omega(2^n/n)$ gates to implement, i.e. most n-bit Boolean functions are "hard". By dimension-arguments as above, we can make analogous statements about unitaries.

One can then ask a complexity-theoretic question: Is the hardness of implementing an arbitrary unitary "equivalent" to the hardness of implementing an arbitrary boolean function? Framed more rigorously, suppose $f : \{0,1\}^{\text{poly}(n)} \to \{0,1\}$ is a Boolean function on $\text{poly}(n)$-bits. Let A^f be an oracle for f, i.e. given any $x \in \{0,1\}^{\text{poly}(n)}$, A_f can, in one query, give us the value of $f(x)$. Then the **Unitary Synthesis Problem** asks:

Problem. Given any n-qubit unitary U, can we implement/approximate U using a $\text{poly}(n)$ sized quantum circuit, given that we're allowed to invoke A_f, for any $f : \{0,1\}^{\text{poly}(n)} \to \{0,1\}$?

Most believe that the answer to the Unitary Synthesis Problem is no, i.e. there exist n-qubit unitaries that are hard to implement/approximate even if we're allowed access to A_f. Stated differently, the hardest quantum problems are harder than the hardest classical problems.

While we don't have an answer to the Unitary Synthesis Problem yet, Lombardi, Ma and Wright [LMW24] proved that if we're only allowed to invoke A_f once, then there exist n-qubit unitaries that can't be approximated using any $\text{poly}(n)$-sized quantum circuit.

[7] when we say "approximate", we mean compute a matrix $\widetilde{U}$ using a $\text{poly}(n)$-sized circuit such that $\|U - \widetilde{U}\| \leq \exp(-\Omega(n))$, where $\|\cdot\|$ is the operator norm

[8] in the topology induced by the operator norm

§2. Quantum Information Theory

2.1. Superdense Coding

Shannon's information theory tells us that one can't communicate more than n bits of information by sending only n bits.
We shall see a quantum violation of this fact, by showing that one can communicate two bits of information by sending only one qubit over. This protocol requires entanglement though. Without entanglement, **Holevo's theorem** tells us that it is impossible to communicate more than 1 bit of information by transmitting only 1 qubit.
Denote by X, Y, Z the *Pauli gates*, where:

Definition 2.1 (Pauli Gates)**.** We define:

$$X := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, Y := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, Z := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Note that X is just the NOT gate.

Now, suppose we have the EPR pair $|\psi\rangle := \frac{|00\rangle+|11\rangle}{\sqrt{2}}$. Then applying the X, Z gates to the first qubit yields:

$$(X \otimes I)|\psi\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}} =: |\psi_1\rangle$$

$$(Z \otimes I)|\psi\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}} =: |\psi_2\rangle$$

$$(Z \otimes I)(X \otimes I)|\psi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}} =: |\psi_3\rangle$$

Note that $V := \{|\psi\rangle, |\psi_1\rangle, |\psi_2\rangle, |\psi_3\rangle\}$ form an orthonormal basis (of $\mathbb{C}^4$).
Now suppose Alice has one qubit of the EPR pair, and Bob has the other. Also suppose Alice wants to transmit two bits $x, y \in \{0, 1\}$ to Bob. So she:

1. Applies the X gate to her qubit if $x = 1$.
2. Applies the Z gate to her qubit if $y = 1$.

Alice then sends her qubit to Bob. Bob, who now has two (entangled) qubits, applies the following transform to it:

$$U := \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

The key point of U is that $U|\psi\rangle = |0\rangle \otimes |0\rangle, U|\psi_1\rangle = |1\rangle \otimes |0\rangle, U|\psi_2\rangle = |0\rangle \otimes |1\rangle, U|\psi_3\rangle = |1\rangle \otimes |1\rangle$, i.e. U "decodes" the coding procedure of Alice. Also note that U is just a basis changing matrix, which changes the basis V for $\mathbb{C}^4$ to the canonical basis $\{|0\rangle \otimes |0\rangle, |1\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |1\rangle\}$ for $\mathbb{C}^4$.
Thus by applying U and then measuring in the usual $\{|0\rangle, |1\rangle\}^{\otimes 2}$ basis, Bob can uniquely recover the bits x, y.
One might wonder if we can push this further, i.e. use 1 qubit to transmit more than 2 bits of information (with any number of pre-shared entangled qubits). However, it can be shown that that is impossible. Thus, to summarize,

Theorem 2.1 (Superdense Coding)**.** Through entanglement, one can use a qubit to transmit ≤ 2 bits of information. Without entanglement, a qubit can transmit ≤ 1 bit of information. More precisely, if we denote by 'ebits' the number of pre-shared entangled bits, then we have: 1 qubit + 1 ebit ≤ 2 bits, 1 qubit + any number of ebits ≤ 2 bits, 1 qubit ≤ 1 bit.

What about the converse? What about if we want to send a qubit over? We shall now describe a protocol which can be summarized as 2 bits $+ 1$ ebit $= 1$ qubit. This "equation" is also optimal, i.e. in general we have 2 bits $+ 1$ ebit ≤ 1 qubit.
The problem is as follows: Alice has the qubit with state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. She has to transmit it to Bob using only classical channels.
To do so, she has one qubit from an EPR pair $\frac{|00\rangle+|11\rangle}{\sqrt{2}}$, where the other pair is with Bob. Alice applies the transform $(H \otimes I) \cdot \mathrm{CNOT}$ to her qubits, where her first qubit is $|\psi\rangle$, and her second qubit is the one from the EPR pair.
At the end of this transformation, she measures her qubits, to obtain $x, y \in \{0, 1\}$. She then sends x, y to Bob.
Let's analyze this protocol. The joint state of the 3 qubits (two with Alice, one with Bob) is

$$\big(\alpha|0\rangle + \beta|1\rangle\big) \otimes \frac{|00\rangle + |11\rangle}{\sqrt{2}} = \frac{\alpha|000\rangle + \alpha|011\rangle + \beta|100\rangle + \beta|111\rangle}{\sqrt{2}}$$

When we apply CNOT to the first two qubits (which belong to Alice), we get the state

$$\frac{\alpha|000\rangle + \alpha|011\rangle + \beta|110\rangle + \beta|101\rangle}{\sqrt{2}}$$

Finally, applying H to the first qubit yields

$$\frac{\alpha|{+}00\rangle + \alpha|{+}11\rangle + \beta|{-}10\rangle + \beta|{-}01\rangle}{\sqrt{2}} = \frac{\alpha|000\rangle + \alpha|100\rangle + \alpha|011\rangle + \alpha|111\rangle + \beta|010\rangle - \beta|110\rangle + \beta|001\rangle - \beta|101\rangle}{2}$$

$$= \frac{1}{2}\left[|00\rangle \otimes (\alpha|0\rangle + \beta|1\rangle) + |01\rangle \otimes (\beta|0\rangle + \alpha|1\rangle) + |10\rangle \otimes (\alpha|0\rangle - \beta|1\rangle) + |11\rangle \otimes (-\beta|0\rangle + \alpha|1\rangle)\right]$$

Thus, if Alice reports 00 to Bob, Bob already has the qubit. If she reports 01, Bob applies the $\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} = \mathrm{NOT}$ gate, if she reports 10, Bob applies the $\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$ gate, and if she reports 11, Bob applies the $\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix} = \mathrm{NOT}$ gate.
Thus Bob has the qubit with state $|\psi\rangle$. Also note that Alice destroys her own copy through measurement, which agrees with the No-Cloning theorem, because otherwise the state $|\psi\rangle$ would have been cloned.
This protocol can be generalized to transport any arbitrary n-qubit state $|\psi\rangle$ using n ebits and $2n$ classical bits.

2.2. Quantifying Entanglement

Consider the *GHZ state*, given by $\frac{|000\rangle+|111\rangle}{\sqrt{2}}$. Clearly the 3 qubits of the GHZ state are highly entangled with each other. Now, consider the Reduced Density Matrix of the first two bits. It is

$$\frac{|00\rangle}{\sqrt{2}} \cdot \frac{\langle 00|}{\sqrt{2}} + \frac{|11\rangle}{\sqrt{2}} \cdot \frac{\langle 11|}{\sqrt{2}} = \frac{1}{2}\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

Note that all the off-diagonal entries are 0, i.e. the first two qubits are not entangled with each other as a subsystem, even though as a whole, they are highly entangled. This phenomenon is also known as *monogamy of entanglement*: Indeed, if ρ_{AB} is a maximally mixed state which is also a reduced state of the system ABC, then $\rho_{ABC} = \rho_{AB} \otimes \rho_C$, i.e. once A and B get entangled maximally, they have "no more entanglement left" for C. Conversely, since the GHZ state is highly (maximally) entangled, its constituents themselves have no entanglement with each other.
Thus, to answer questions about how much entanglement is possible in a system, we introduce the notion of von Neumann entropy:

Definition 2.2 (von Neumann Entropy)**.** Let ρ be the density matrix of some quantum system. Let $\{\gamma_i\}$ be the eigenvalues of ρ. Then the entropy of ρ is defined to be:

$$S(\rho) := -\sum \gamma_i \log_2 \gamma_i$$

We set $0 \cdot \lg 0$ to be 0.

Remark. A few remarks are due:

1. **Upper Bound**: Using the concavity of $-x \log_2 x$, one can show that if $\rho \in \mathbb{C}^{N \times N}$, then $S(\rho) \leq \log_2 N$.
2. **Independence**: $S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B)$.
3. **Concavity**: S is a concave function, i.e. if $\rho_1, \ldots, \rho_n$ are density matrices, then $S(\sum_{i=1}^n \lambda_i \rho_i) \geq \sum_{i=1}^n \lambda_i S(\rho_i)$ for $\lambda_1, \ldots, \lambda_n \in [0,1]$ such that $\sum_{i=1}^n \lambda_i = 1$.
4. **Reverse Concavity**: We also have $S(\sum_{i=1}^n \lambda_i \rho_i) \leq \sum_{i=1}^n \lambda_i S(\rho_i) + H(\{\lambda_i\})$, where $H(\{\lambda_i\}) = -\sum \lambda_i \lg \lambda_i$. Equality is achieved if ρ_i have orthogonal support, i.e. if V_i is the subspace of $\mathcal{H}$ spanned by the eigenvectors of the non-zero eigenvalues of ρ_i, then $V_i \subseteq V_j^{\perp}$ for all $i \neq j$.
5. **Strong Subadditivity**: Let (A, B, C) be a tripartition of a system. Then $S(\rho) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})$. In particular, if we take B to be 0-dimensional, then $S(\rho) \leq S(\rho_A) + S(\rho_C)$.
6. **Triangle Inequality**: Let (A, C) be a partition. Then $|S(\rho_A) - S(\rho_C)| \leq S(\rho)$. We thus have $|S(\rho_A) - S(\rho_C)| \leq S(\rho) \leq S(\rho_A) + S(\rho_C)$ for any partition (A, C).

There is another way to view von Neumann entropy: Suppose ρ is some state, possibly mixed. Note that measuring ρ against some basis $V := \{|\psi_1\rangle, \ldots, |\psi_N\rangle\}$ yields a probability distribution on V, namely, where the probability associated to $|\psi_i\rangle$ is the probability that it is the outcome of the measurement.
Furthermore, to any probability distribution $\{p_i\}$, we can associate a *Shannon entropy*, given as:

$$H\left(\{p_i\}\right) = \sum p_i \log_2 \left(\frac{1}{p_i}\right)$$

Then one can show that

$$S(\rho) = \min_U H(\mathrm{diag}(U \rho U^\dagger))$$

Let $\mathcal{C}$ be the canonical basis in which ρ is expressed. Then $U\mathcal{C} = \{|\psi_1\rangle, \ldots, |\psi_N\rangle\}$ is an orthonormal basis. Then note that $(U\rho U^\dagger)_{ii} = \Pr(\text{We obtain } |\psi_i\rangle \text{ on measuring } \rho \text{ against } U\mathcal{C})$. Thus $\mathrm{diag}(U\rho U^\dagger)$ contains the probabilities of obtaining individual basis vectors in $U\mathcal{C}$, and $H(\cdot)$ represents the entropy of this distribution.
Note that the above characterization immediately shows that the von Neumann entropy of pure states is 0: Indeed, if we measure the density matrix of a pure state against a basis containing the state, we get a deterministic outcome, i.e. a probability distribution with support of size 1, which has 0 entropy. Another way of seeing this is: If ρ is a pure state, then ρ has exactly one non-zero eigenvalue, which must be 1, since $\mathrm{tr}(\rho) = 1$. But $\log(1) = 0$, and thus $S(\rho) = 0$.
Before we can quantify the entanglement of a system, we introduce the notion of *Schmidt decomposition*: Let $|\psi\rangle = \sum_{i,j} \alpha_{ij} |i\rangle_A \otimes |j\rangle_B$ be a pure state. Write $A := 2^a, B := 2^b$, and WLOG assume $A \geq B$. Let $\alpha \in \mathbb{C}^{A \times B}$ be the matrix collecting the coefficients α_{ij}. The SVD decomposition of α looks something like:

$$\alpha = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^\dagger$$

where $U \in \mathbb{C}^{A \times A}, V \in \mathbb{C}^{B \times B}$ are unitary matrices, and $\Sigma \in \mathbb{C}^{B \times B}$ is a square diagonal matrix with non-negative entries. Write $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$, where $U_1 \in \mathbb{C}^{A \times B}$. Then $\alpha = U_1 \Sigma V^\dagger$. Finally, writing $U = \begin{bmatrix} |u_1\rangle & \cdots & |u_B\rangle \end{bmatrix}, V =$

$\begin{bmatrix}|v_1\rangle & \cdots & |v_B\rangle\end{bmatrix}, \Sigma = \begin{bmatrix}\lambda_1^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_B^2\end{bmatrix}$, we get that $\alpha = \sum_{\iota=1}^{B} \lambda_\iota^2 |u_\iota\rangle\langle v_\iota|$.
Now, note that there is a linear bijective correspondence between $\{|i\rangle_A \otimes |j\rangle_B\}$ and $\{|i\rangle_A\langle j|_B\}$. Thus,

$$|\psi\rangle = \sum_{i,j,\iota} (\lambda_\iota^2 |u_\iota\rangle\langle v_\iota|)_{ij} |i\rangle_A \otimes |j\rangle_B = \sum_\iota \lambda_\iota^2 u_\iota \otimes v_\iota^*$$

Let r be the number of strictly positive λ_*. Then r is known as the *Schmidt rank* of $|\psi\rangle$.
We can now define the *entanglement entropy* of a *pure state*:

Definition 2.3. Let ρ be the density matrix of a *pure state*. Then the entanglement entropy of ρ w.r.t the partition (A, B) is defined to be $S(\rho_A) = S(\rho_B)$.

Remark. A few remarks are due:

1. The pure state assumption is necessary for the above definition. Mixed states make life complicated since von Neumann entropy is not linear.
2. Why is $S(\rho_A) = S(\rho_B)$? Let $\alpha \in \mathbb{C}^{2^a \times 2^b}$ be the matrix collecting the coefficients α_{ij}. Then $\rho_B = (\alpha^\dagger \alpha)^\mathsf{T} = \alpha^\mathsf{T} \alpha^*$. On the other hand, for the computation of ρ_A, we have to replace α by α^T, and we get $\rho_A = \alpha\alpha^\dagger$. Note that $\alpha\alpha^\dagger$ and $\alpha^\mathsf{T}\alpha^*$ have the same non-zero spectrum, as can easily be seen by writing the SVD decomposition of α. Since the definition of von Neumann entropy only depends on the non-zero spectrum, $S(\rho_A) = S(\rho_B)$.
3. Note that the notion of entanglement entropy is basis independent, as it should be.
4. Let $|\psi\rangle$ be a pure state, and let r_ψ be its Schmidt rank. If $r_\psi = 1$, then $|\psi\rangle = |u_A\rangle \otimes |v_B\rangle$ for some $|u_A\rangle, |v_B\rangle$. Then note that $\rho_A = |u_A\rangle\langle u_A|$, i.e. ρ_A is a pure state. Consequently, $S(\rho_A) = 0$, i.e. the entanglement entropy of $|\psi\rangle$ w.r.t the partition (A, B) is 0! Conversely, suppose $r_\psi > 1$. Note that the non-zero eigenvalues of $\alpha\alpha^\dagger$ are precisely $\lambda_1^2, \ldots, \lambda_{r_\psi}^2$, by the properties of the SVD decomposition. Since $r_\psi > 1$, $S(\rho_A) = H\left(\left\{\lambda_1^2, \ldots, \lambda_{r_\psi}^2\right\}\right) > 0$, since the entropy of any probability distribution with support size > 1 is non-zero. Summarizing the entire discussion above, we have:

   Theorem 2.2. Let $|\psi\rangle = \sum_{i,j} \alpha_{i,j} |i\rangle_A \otimes |j\rangle_B$ be a pure state, where $\{|i\rangle_A\}$, $\{|j\rangle_B\}$ are arbitrary bases of $\mathcal{H}_A, \mathcal{H}_B$ respectively. The number of non-zero eigenvalues of $\alpha\alpha^\dagger$ is known as the Schmidt rank of $|\psi\rangle$. If the Schmidt rank of $|\psi\rangle$ is 1, then $|\psi\rangle$ has 0 entanglement entropy. Otherwise, the entanglement entropy of $|\psi\rangle$ is given by $H(\{\lambda_*^2\})$, where $\{\lambda_*^2\}$ are the non-zero eigenvalues of $\alpha\alpha^\dagger$.

5. Note that the entanglement entropy of a state depends on the partition. For example, take $|\psi\rangle = |0\rangle \otimes \frac{|00\rangle + |11\rangle}{\sqrt{2}}$. Let's label the qubits $1, 2, 3$. Then the entanglement entropy of $|\psi\rangle$ w.r.t the partition $\{\{1\}, \{2, 3\}\}$ is 0. However, the entanglement entropy w.r.t the partition $\{\{1, 2\}, \{3\}\}$ is not zero.

Example. Let's see a few examples of entanglement entropy:

1. Consider the EPR state $\frac{|00\rangle + |11\rangle}{\sqrt{2}}$. The Reduced Density Matrix of Bob is $\frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$, whose eigenvalues are $\frac{1}{2}, \frac{1}{2}$. Thus the entanglement entropy of the EPR state w.r.t. the natural partition is $\frac{1}{2} \cdot \log_2(2) + \frac{1}{2} \cdot \log_2(2) = 1$. Note that 1 is the maximum possible entropy for 2×2 matrices, and thus the EPR state has the maximum possible entropy.

2. Consider $|\psi\rangle = \frac{3}{5}|0\rangle|+\rangle + \frac{4}{5}|1\rangle|-\rangle$. Then the Reduced Density Matrix for Bob is

$$\frac{9}{25}|+\rangle\langle+| + \frac{16}{25}|-\rangle\langle-|$$

Note that the eigenvalues of a matrix remain the same regardless of what basis we express it in. Thus, expressing the above matrix in $\{|+\rangle, |-\rangle\}$ basis, we get the matrix $\begin{bmatrix} 9/25 & 0 \\ 0 & 16/25 \end{bmatrix}_{\{|+\rangle,|-\rangle\}}$, whose eigenvalues are $\frac{9}{25}, \frac{16}{25}$. Consequently, the entanglement entropy is

$$\frac{9}{25}\log_2\frac{25}{9} + \frac{16}{25}\log_2\frac{25}{16} \approx 0.942$$

What this means is that if Alice and Bob shared 1000 copies of $|\psi\rangle$ among themselves, then no protocol could transmit more than 942 qubits by exploiting the entanglement of $|\psi\rangle$.

2.2.1. Quantifying Entanglement for Mixed States

Suppose we're in the quantum teleportation setting. Also assume that Alice and Bob can communicate as much as they want via classical channels, and both of them can do as many local operations on their qubits as they want. This assumption is also known as unbounded 'LOCC' (Local Operations and Classical Communication).
Let ρ be a density matrix. Define $E_F(\rho)$ to be the number of ebits needed to teleport ρ under the assumptions of unbounded LOCC, and the fact that Alice and Bob are using the best possible protocol they can. The 'F' in E_F stands for 'Formation'. Define $E_D(\rho)$ to be the number of ebits extracted from ρ under the assumptions of unbounded LOCC, and the fact that Alice and Bob are using the best possible protocol they can. The 'D' in E_D stands for 'Distillation'. Firstly, $E_F \geq E_D$: If we could create more ebits out of ρ than it took to make ρ, then we would have an infinite money glitch, which can't occur. Secondly, it can be shown that if ρ is a pure state, then $E_F = E_D =$ The Entanglement Entropy of ρ. However, if ρ is a mixed state, then it is even possible that $E_F > 0 = E_D$, i.e. no entanglement can be distilled from ρ, even though ρ is entangled. Such a ρ is called a *bound entangled state*.
We call a state ρ *separable* if $\rho = \sum p_i|u_i\rangle_A\langle u_i|_A \otimes |v_i\rangle_B\langle v_i|_B$ for some $\{|u_i\rangle_A\} \in \mathcal{H}_A, \{|v_i\rangle_B\} \in \mathcal{H}_B$, and probability distribution $\{p_i\}$. A state is called entangled if it is not separable.
Note that if ρ is pure, then the Schmidt rank criterion gives us a way to decide if ρ is entangled or separable. However, if ρ is mixed, then a result due to Gurvits says that deciding if ρ is entangled or separable is NP-hard!

2.3. Bell's Inequality and the CHSH Game

Clauser, Horne, Shimony, and Holt, in 1969, proposed a game called the CHSH game. In the game, we have Alice, Bob, and a referee named Charles. Charles generates two uniformly random bits x and y, and gives them to Alice and Bob. After the game begins, Alice and Bob are not allowed to communicate; However, they can decide on some strategy beforehand though.
On receiving x, Alice must propose to Charles some $a = a(x) \in \{0, 1\}$. Similarly, on receiving y, Bob must propose to Charles some $b = b(y) \in \{0, 1\}$. Alice and Bob are said to have won the game if $a(x) +_{\mathbb{F}_2} b(y) = xy$.
Note that there are 4 possible boolean functions $a(\cdot) : \{0, 1\} \to \{0, 1\}$: $a(\cdot) = 0, a(\cdot) = 1, a(\cdot) = \cdot, a(\cdot) = 1 - \cdot$, and thus there are 16 possible strategies for Alice and Bob combined.

Alice/Bob	$b(\cdot)=0$			$b(\cdot)=1$			$b(\cdot)=\cdot$			$b(\cdot)=1-\cdot$		
$a(\cdot)=0$	x	y	Outcome	x	y	Outcome	x	y	Outcome	x	y	Outcome
	0	0	W	0	0	L	0	0	W	0	0	L
	0	1	W	0	1	L	0	1	L	0	1	W
	1	0	W	1	0	L	1	0	W	1	0	L
	1	1	L	1	1	W	1	1	W	1	1	L
$a(\cdot)=1$	x	y	Outcome	x	y	Outcome	x	y	Outcome	x	y	Outcome
	0	0	L	0	0	W	0	0	L	0	0	W
	0	1	L	0	1	W	0	1	W	0	1	L
	1	0	L	1	0	W	1	0	L	1	0	W
	1	1	W	1	1	L	1	1	L	1	1	W
$a(\cdot)=\cdot$	x	y	Outcome	x	y	Outcome	x	y	Outcome	x	y	Outcome
	0	0	W	0	0	L	0	0	W	0	0	L
	0	1	W	0	1	L	0	1	L	0	1	W
	1	0	L	1	0	W	1	0	L	1	0	W
	1	1	W	1	1	L	1	1	L	1	1	W
$a(\cdot)=1-\cdot$	x	y	Outcome	x	y	Outcome	x	y	Outcome	x	y	Outcome
	0	0	L	0	0	W	0	0	L	0	0	W
	0	1	L	0	1	W	0	1	W	0	1	L
	1	0	W	1	0	L	1	0	W	1	0	L
	1	1	L	1	1	W	1	1	W	1	1	L

The above table collects the results of all possible strategies Alice and Bob could adopt. 'W' stands for 'Win', while 'L' stands for 'Lose'. Thus, no deterministic strategy can make Alice and Bob win more than 75% of the time (this fact is known as *Bell's Inequality*). Furthermore, since any randomized strategy on Alice and Bob's part will just be a mixing of some of the above 16 deterministic strategies, it follows that no randomized strategy can make Alice and Bob win more than 75% of the time either. Thus, no classical strategy can make Alice and Bob win more than 75% of the time.

We shall now see how entanglement can make Alice and Bob win *more* than 75% of the time.

Define $v_\theta := \cos(\theta)|0\rangle + \sin(\theta)|1\rangle$. Also suppose Alice and Bob have a qubit each from an EPR pair (which is also known as a Bell pair). Then Alice and Bob's strategy is:

1. If $x = 0$, Alice measures her qubit in the $\{|0\rangle, |1\rangle\}$ basis. If she obtains $|0\rangle$, she sends 0 to Charles. If she obtains $|1\rangle$, she sends 1 to Charles.
2. If $x = 1$, Alice measures her qubit in the $\{|+\rangle, |-\rangle\}$ basis. If she obtains $|+\rangle$, she sends 0 to Charles. If she obtains $|-\rangle$, she sends 1 to Charles.
3. If $y = 0$, Bob measures his qubit in the $\{v_{\pi/8}, v_{5\pi/8}\}$ basis. If he obtains $v_{\pi/8}$, he sends 0 to Charles. If he obtains $v_{5\pi/8}$, he sends 1 to Charles.
4. If $y = 1$, Bob measures his qubit in the $\{v_{-\pi/8}, v_{3\pi/8}\}$ basis. If he obtains $v_{-\pi/8}$, he sends 0 to Charles. If he obtains $v_{3\pi/8}$, he sends 1 to Charles.

How do we analyze the winning probability of this protocol? WLOG assume Alice measures first. We have 4 cases:

1. $x = y = 0$: Note that Alice and Charles lose only if they send different bits to Charles, which can only happen if Alice measures $|0\rangle$ and Bob measures $v_{5\pi/8}$, or if Alice measures $|1\rangle$ and Bob measures $v_{\pi/8}$. If Alice measures $|0\rangle$, then Bob's qubit collapses to $|0\rangle$. The probability that he obtains $v_{5\pi/8}$ is $|\langle v_{5\pi/8}|0\rangle|^2 = \cos^2(5\pi/8) = \sin^2(\pi/8)$. If Alice gets $|1\rangle$, the probability that Bob gets $v_{\pi/8}$ is $\sin^2(\pi/8)$. Thus, the probability that Alice and Bob lose is $\sin^2(\pi/8)$, or alternatively, their winning probability is $1 - \sin^2(\pi/8) = \cos^2(\pi/8)$.
2. The cases of $x = 0, y = 1$, and $x = 1, y = 0$ are similar. In both of them the winning probability is $\cos^2(\pi/8)$.
3. $x = y = 1$: Alice and Charles lose only if Alice measures $|+\rangle$ and Bob measures $v_{3\pi/8}$, or if Alice measures $|-\rangle$ and Bob measures $v_{-\pi/8}$. Note that Alice measuring in the $\{|+\rangle, |-\rangle\}$ basis is equivalent to Hadamarding her

qubit and then measuring in the original basis. Now, $(H \otimes I) \cdot \frac{|00\rangle+|11\rangle}{\sqrt{2}} = \frac{|00\rangle+|01\rangle+|10\rangle-|11\rangle}{2}$. Getting $|+\rangle$ is now equivalent to getting $|0\rangle$. If Alice gets $|0\rangle$, Bob's state collapses to $|+\rangle$. The probability that Bob gets $v_{3\pi/8}$ now is $\left|\left\langle v_{3\pi/8}|+\right\rangle\right|^2 = \sin^2(\pi/8)$. If Alice gets $|1\rangle$, Bob's state collapses to $|-\rangle$. The probability that Bob gets $v_{-\pi/8}$ now is $\left|\left\langle v_{-\pi/8}|-\right\rangle\right|^2 = \sin^2(\pi/8)$. Thus, once again the winning probability is $\cos^2(\pi/8)$.

Thus, using entanglement, Alice and Bob can win $\cos^2(\pi/8) \approx 85\%$ of the time, which is significantly more than 75%. Furthermore, while $\cos^2(\pi/8)$ might seem like an artefact of the above protocol, **Tsirelson's bound** shows that *no quantum protocol* can win in the CHSH game with probability more than $\cos^2(\pi/8)$. [9]

A philosophical point: Note that in *classical local realism*, we assume that Alice and Bob can't communicate faster than the speed of light, and thus in the context of the CHSH experiment, they can't communicate at all. We also assume that our universe is classical. However, in the classical setting, we can't beat 75% in the CHSH game. Since experiments have repeatedly demonstrated that we can beat 75% in CHSH, our universe doesn't follow classical local realism. But note that to beat 75% in a purely classical universe, we would have to have communication between Bob and Alice in the game, which would entail faster-than-light communication. However, our universe doesn't support faster-than-light communication either. The only conclusion is that our universe is fundamentally quantum, and that any classical "simulation" of our universe would necessarily have faster-than-light communication.

[9] while Tsirelson's bound is somewhat tricky to prove, the following weaker version is much easier to establish: Suppose Alice measures in the $\{v_{\theta_0}, v_{\theta_0+\pi/2}\}$ basis if $x = 0$, and measures in the $\{v_{\theta_1}, v_{\theta_1+\pi/2}\}$ basis if $x = 1$. Similarly, Bob measures in the $\{v_{\phi_0}, v_{\phi_0+\pi/2}\}$ basis if $y = 0$, and measures in the $\{v_{\phi_1}, v_{\phi_1+\pi/2}\}$ basis if $y = 1$. Then these class of protocols can't achieve more than $\cos^2(\pi/8)$.

§3. Quantum Algorithms

Given an arbitrary unitary U, constructing the smallest circuit to simulate/approximate U is a very difficult problem. Thus, to sidestep this difficulty, we study a more restricted model of computation, called query complexity. In this, we're given a boolean function $f : \{0,1\}^n \to \{0,1\}$. We are also given an oracle $\mathcal{O}_f$ which acts as $\mathcal{O}_f(|x\rangle) = (-1)^{f(x)}|x\rangle$. The only thing we care about is how many calls we have to make to $\mathcal{O}_f$ to achieve some particular task. Furthermore, in practice, querying $\mathcal{O}_f$ is often the most time-consuming step. Consequently, the query complexity is also a very good proxy for the running time of the algorithm.

The oracle $\mathcal{O}_f$ might seem weird at first. A more natural thing to do would be to send $|x\rangle$ to $|f(x)\rangle$. However, this operation is not reversible. The next best thing to do would be to add an ancilla bit, and perform a CNOT type operation, i.e. $|x\rangle \otimes |b\rangle \mapsto |x\rangle \otimes |b \oplus f(x)\rangle$, and in particular $|x\rangle \otimes |0\rangle \mapsto |x\rangle \otimes |f(x)\rangle$. This oracle is known as the XOR-oracle. As it turns out, the XOR-oracle is equivalent to $\mathcal{O}_f$: Indeed, suppose we had the XOR-oracle, and we want to simulate $\mathcal{O}_f$. Given the input $|x\rangle$ to $\mathcal{O}_f$, pass $|x\rangle \otimes |-\rangle$ to the XOR-oracle. The XOR-oracle converts it to $\frac{|x\rangle\otimes|f(x)\rangle - |x\rangle\otimes|1-f(x)\rangle}{\sqrt{2}}$. If $f(x) = 0$, $\frac{|x\rangle\otimes|f(x)\rangle - |x\rangle\otimes|1-f(x)\rangle}{\sqrt{2}} = |x\rangle \otimes |-\rangle$. If $f(x) = 1$, $\frac{|x\rangle\otimes|f(x)\rangle - |x\rangle\otimes|1-f(x)\rangle}{\sqrt{2}} = -|x\rangle \otimes |-\rangle$. Thus $\mathsf{XOR}(|x\rangle \otimes |-\rangle) = (-1)^{f(x)}|x\rangle \otimes |-\rangle$. Note that we have managed to implement $\mathcal{O}_f$.

3.1. Deutsch-Jozsa Algorithm [DJ92]

Suppose we have a function $f : \{0,1\}^n \to \{0,1\}$, and we're promised that f is either a constant function, or is *balanced*, i.e. $|f^{-1}(1)| = 2^{n-1}$. We have to determine which of the two cases it is, with *perfect correctness*, [10] i.e. we have to get the correct answer with probability 1. It is easy to see that any classical algorithm for this must necessarily query $f \geq 2^{n-1} + 1$ times. We present the Deutsch-Jozsa algorithm which achieves this just 1 query to $\mathcal{O}_f$!
The circuit which achieves this is as follows: $H^{\otimes n} \cdot \mathcal{O}_f \cdot H^{\otimes n}$. Suppose we apply this circuit to $|0\rangle^{\otimes n}$. On application of $H^{\otimes n}$, our state becomes

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

Applying $\mathcal{O}_f$ to this state makes it $\frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)}|x\rangle$. Now, when we apply $H^{\otimes n}$ on $|x\rangle$, we obtain the state

$$\bigotimes_{i=1}^{n} \frac{|0\rangle + (-1)^{x_i}|1\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2^n}} \sum_{y\in\{0,1\}^n} (-1)^{\langle x,y\rangle}|y\rangle$$

Thus, our final state is

$$|\psi\rangle := \frac{1}{2^n} \sum_{x,y\in\{0,1\}^n} (-1)^{f(x)+\langle x,y\rangle}|y\rangle$$

In particular, the coefficient of $|0\rangle^{\otimes n}$ in the above sum is

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

Note that if f is constant, then the above term is ± 1, while if f is balanced then the above term is 0, i.e. when we measure $|\psi\rangle$, either we'll obtain $|0\rangle^{\otimes n}$ with probability 1, or we'll not obtain it at all! Thus, when we measure the state $|\psi\rangle$, if we obtain $|0\rangle^{\otimes n}$, then f is constant, otherwise it is balanced.

[10] If we wanted a classical algorithm that succeeds with probability $1-\delta$, then $\mathcal{O}(\log 1/\delta)$ queries to f suffices. Thus, for all "practical purposes", the above quantum speedup is not all that impressive

3.2. Bernstein-Vazirani Algorithm [BV97]

Suppose we are given $f : \{0,1\}^n \to \{0,1\}$, and we're promised that $f(x) = \langle s, x\rangle$ for some $s \in \{0,1\}^n$. Our task is to find s.
Note that any classical algorithm for this must make $\geq n$ queries to f (because otherwise the system of linear equations will remain under-determined). On the other hand, 1 query to $\mathcal{O}_f$ suffices!
Turns out the exact same circuit as Deutsch-Jozsa works! Thus apply $H^{\otimes n} \cdot \mathcal{O}_f \cdot H^{\otimes n}$ to $|0\rangle^{\otimes n}$. We get the state

$$\frac{1}{2^n} \sum_{x,y\in\{0,1\}^n} (-1)^{f(x)+\langle x,y\rangle} |y\rangle = \frac{1}{2^n} \sum_{x,y\in\{0,1\}^n} (-1)^{\langle x,s\rangle+\langle x,y\rangle} |y\rangle$$

The coefficient of $|s\rangle$ in the above state is $\frac{1}{2^n}\sum_{x\in\{0,1\}^n}(-1)^{\langle x,s\rangle+\langle x,s\rangle} = 1$. Thus, simply measuring $|\psi\rangle$ gets us the string s!

3.3. Simon's Algorithm [Sim97]

We are given a function $f : \{0,1\}^n \to \{0,1\}^n$. The function satisfies the property that there exists a string $s \in \{0,1\}^n, s \neq 0^n$, such that

$$f(x) = f(y) \iff x = y \oplus s, \forall x, y \in \{0,1\}^n, x \neq y$$

Note that the above condition implies that any element in $f\left(\{0,1\}^n\right)$ has exactly two pre-images in $\{0,1\}^n$. Furthermore, for any $z \in f\left(\{0,1\}^n\right)$, the two pre-images of z xor upto s. Thus, it suffices to find a *collision* in f to calculate s, i.e. find two distinct $x, y \in \{0,1\}^n$ such that $f(x) = f(y)$. Then $x \oplus y$ gives us s.
Again, any deterministic classical algorithm must make $\geq 2^{n-1} + 1$ queries to f to compute s. What about randomized classical algorithms? Using a standard birthday paradox argument, to find (with high probability) a collision among the 2^n input values of f, it suffices to sample $\mathcal{O}^*(\sqrt{2^n})$ [11] uniformly random strings in $\{0,1\}^n$ and query f on them.
Using a so-called *adversary argument*, one can show that *any* randomized classical algorithm must also take $\Omega^*(\sqrt{2^n})$ [12] queries: To do so, we invoke *Yao's Minimax principle*:

Yao's Minimax Principle

Suppose we have a problem with inputs coming from the set $\mathcal{X}$. Let $\mathcal{A}$ be the set of all deterministic algorithms solving the problem. For any $a \in \mathcal{A}, x \in \mathcal{X}$, let $c(a, x)$ be the runtime of the algorithm a on the input x.
Let π be *any* probability distribution on $\mathcal{A}$, and let σ be *any* probability distribution on $\mathcal{X}$. Then

$$\max_{x\in\mathcal{X}} \mathop{\mathbb{E}}_{a\sim\pi} c(a, x) \geq \min_{a\in\mathcal{A}} \mathop{\mathbb{E}}_{x\sim\sigma} c(a, x)$$

Now, let a_0 be our randomized algorithm. Note that a_0 naturally corresponds to a distribution π on $\mathcal{A}$, where $\mathcal{A}$ is the set of all deterministic algorithms for our problem. Thus $\mathbb{E}_{a\sim\pi} c(a, \cdot) = c(a_0, \cdot)$. Also note that for our problem, the "set of inputs" $\mathcal{X}$ is just the set of possible 's', i.e. $\mathcal{X} = \{0,1\}^n \setminus 0^n$. Let σ be the uniform distribution on $\mathcal{X}$, i.e. s is chosen uniformly amongst $\{0,1\}^n \setminus 0^n$. Also, let $a_* \in \operatorname{argmin}_{a\in\mathcal{A}} \mathbb{E}_{s\sim\sigma} c(a, s)$. Let the series of queries in a_* be $x_0, x_1, \ldots$. Now, suppose we have queried t values. Then we have managed to generate $\leq \binom{t}{2}$ possible candidates for s, and thus the probability that we have found a collision is $\leq \frac{\binom{t}{2}}{2^n-1}$.
Thus, by the union bound, we won't observe any collisions until $t = \Omega^*(\sqrt{2^n})$, and thus $\mathbb{E}_{s\sim\sigma} c(a_*, s) = \Omega^*(\sqrt{2^n})$. But

$$\max_{s\in\{0,1\}^n\setminus 0^n} c(a_0, s) \geq \min_{a\in\mathcal{A}} \mathop{\mathbb{E}}_{s\sim\sigma} c(a, s) = \mathop{\mathbb{E}}_{s\sim\sigma} c(a_*, s) = \Omega^*(\sqrt{2^n})$$

Thus, for any randomized algorithm, picking out the worst-case s leads to a run-time of $\Omega^*(\sqrt{2^n})$, as desired.
We will now show that a quantum algorithm named *Simon's algorithm* can perform the task with $\mathcal{O}(n)$ queries to $\mathcal{O}_f$. In this context, we define:

$$\mathcal{O}_f(|x\rangle \otimes |0\rangle^{\otimes n}) := |x\rangle \otimes |f(x)\rangle$$

[11] $\mathcal{O}^*(f(n))$ stands for $\mathcal{O}(f(n) \cdot \text{poly}(n))$
[12] $\Omega^*(f(n))$ means $\Omega^*(f(n)/\text{poly}(n))$

Simon's algorithm goes as follows: We begin with $|0\rangle^{\otimes 2n}$. The first n qubits are meant to be the input registers, and the remaining qubits are meant to be the output registers. Apply the circuit $(H^{\otimes n} \otimes I^{\otimes n}) \cdot \mathcal{O}_f \cdot (H^{\otimes n} \otimes I^{\otimes n})$ to our input.
On the application of $(H^{\otimes n} \otimes I^{\otimes n})$, our state becomes

$$\left(\frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle\right) \otimes |0\rangle^{\otimes n}$$

Then the application of $\mathcal{O}_f$ yields

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

The application of $(H^{\otimes n} \otimes I^{\otimes n})$ once again yields:

$$|\psi\rangle := \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \left(\sum_{y \in \{0,1\}^n} (-1)^{\langle x,y \rangle} |y\rangle \right) \otimes |f(x)\rangle = \frac{1}{2^n} \sum_{x,y \in \{0,1\}^n} (-1)^{\langle x,y \rangle} |y\rangle \otimes |f(x)\rangle$$

Write $\mathcal{Z} := f(\{0,1\}^n)$, and for any $z \in \mathcal{Z}$ write $f^{-1}(z) = \{\alpha_z, \beta_z\}$. Then

$$|\psi\rangle = \frac{1}{2^n} \sum_{y \in \{0,1\}^n, z \in \mathcal{Z}} \left((-1)^{\langle \alpha_z, y \rangle} + (-1)^{\langle \beta_z, y \rangle} \right) |y\rangle \otimes |z\rangle = \frac{1}{2^n} \sum_{y \in \{0,1\}^n} |y\rangle \otimes \left(\sum_{z \in \mathcal{Z}} \left((-1)^{\langle \alpha_z, y \rangle} + (-1)^{\langle \beta_z, y \rangle} \right) |z\rangle \right)$$

Now, measure the first n qubits of $|\psi\rangle$. Suppose we get the output $\mathfrak{y}$. Then note that $\langle \alpha_z, \mathfrak{y} \rangle$ and $\langle \beta_z, \mathfrak{y} \rangle$ must have had the same parity for some $z \in \mathcal{Z}$, because otherwise the coefficient of $|\mathfrak{y}\rangle \otimes |\cdot\rangle$ would have been nullified. This is further equivalent to saying that $\langle \alpha_z \oplus \beta_z, \mathfrak{y} \rangle = \langle s, \mathfrak{y} \rangle$ is even. Thus, when we measure the first n qubits and get some output '$\mathfrak{y}$', that tells us $\langle s, \mathfrak{y} \rangle =_{\mathbb{F}_2} 0$. Now, if we repeat Simon's algorithm $k = \mathcal{O}(n)$ times, with high probability we get $(n-1)$ linearly independent outputs (amongst those k) $\mathfrak{y}^{(1)}, \ldots, \mathfrak{y}^{(n)}$ satisfying $\langle \mathfrak{y}^{(i)}, s \rangle = 0 \bmod 2$. At this point, we can run Gaussian elimination (in $\mathbb{F}_2$) to uniquely recover s, as desired.
To make this argument more precise, define $\mathcal{Y} := \{\mathfrak{y} \in \{0,1\}^n : \langle \mathfrak{y}, s \rangle = 0 \bmod 2\}$. Now, note that the probability that some $\mathfrak{y} \in \mathcal{Y}$ gets outputted on measuring the first n qubits of $|\psi\rangle$ is proportional to:

$$\frac{1}{4^n} \sum_{z \in \mathcal{Z}} \left((-1)^{\langle \alpha_z, y \rangle} + (-1)^{\langle \beta_z, y \rangle} \right)^2 = 4^{1-n} \cdot |\mathcal{Z}|$$

In other words, every element of $\mathcal{Y}$ is equally likely to appear as the output of measuring the first n qubits of $|\psi\rangle$. Thus, k runs of Simon's algorithm is equivalent to k uniformly random samples from $\mathcal{Y}$. Now, the probability that the span of the k chosen samples is at most $(n-2)$-dimensional is

$$\leq \sum_{S \subseteq \mathcal{Y}, \dim(S) = n-2} \Pr(\text{span}(\{\mathfrak{y}^{(i)}\}) \subseteq S)$$

Now, note that $\mathcal{Y} = s^\perp$, and thus $\mathcal{Y}$ is a $(n-1)$-dimensional vector space. It has $2^{n-1} - 1$ $(n-2)$-dimensional subspaces. Thus the above quantity becomes

$$(2^{n-1} - 1)\frac{1}{2^k} < 2^{n-k-1}$$

Thus, by choosing $k = 2n$, we can ensure $1 - \exp(-\Omega(n))$ probability of success.
Simon's algorithm also leads to some interesting philosophical points:

1. Doesn't Simon's algorithm prove that there is an exponential gap between quantum and classical algorithms? No. Note that in Simon's algorithm, f is completely black-box. Now, the moment we try to apply Simon's algorithm to some *specific* function f, we also have to account for the possibility that classical algorithms might exploit some particular property of that function to also give a poly-time algorithm. For example, if $A \in \mathbb{F}_2^{n \times n}$

is a $(n-1)$-rank boolean matrix, then $f(x) = Ax$ satisfies the hypotheses of Simon's problem (with s being the non-zero element in the null-space of A). However, in this case, a classical algorithm can simply run Gaussian elimination on A to find s in polynomial time! Till date, no one has been able to find a specific instantiation of a "Simon function" which respects the $\Omega(\sqrt{2}^n)$ lower bound.

2. **Back-Reaction**: Note that in the final step, when we measure, we only measure the first n qubits. This might lead one to think that the remaining n qubits are actually useless, and just an artefact of our oracle. Indeed, one might ask how storing the values of $f(\cdot)$ in the last n qubits affect the first n qubits. This leads to the principle of *back-action* in quantum mechanics, which says that if a system affects another system (in our case, the first n qubits being one system, and the last n qubits being the other system), then the first system itself gets affected.

3.4. Shor's Algorithm [Sho94]

We shall use ideas from Simon's algorithm to factorize numbers. Before doing so, we describe a (classical) reduction from "period-finding" to factorization. Some of the intermediate supporting lemmata has been taken from [Vaz04]. Let N be a natural number. When we say we want an efficient algorithm for factorizing N, we mean an algorithm which runs in time $\mathrm{poly}\log(N)$: Why? Because the number of digits of N is $\mathcal{O}(\log N)$, and we want an algorithm which runs in time polynomial in the number of digits of N. Thus, write $n := \log(N)$. Also, what do we mean by "factorization"? Given N, output a non-trivial factor of N. [13][14]
Now, WLOG N is odd (since otherwise 2 is a non-trivial factor). Now, there exist algorithms [Ber98] which can detect if N is a perfect power in $n^{1+o(1)}$ time. [15] Thus, WLOG assume N is not a perfect power.
Now, randomly choose an integer $2 \leq a < N$. If $\gcd(a, N) > 1$, [16] then $\gcd(a, N)$ is a non-trivial factor of N, and we're done. Thus assume $\gcd(a, N) = 1$. Then there exists a minimal $r \in \mathbb{N}$ such that $a^r \equiv 1 \bmod N$. Now, suppose r is even. Then we have $N \mid (a^r - 1) \iff N \mid (a^{r/2} - 1)(a^{r/2} + 1)$. Note that $N \nmid (a^{r/2} - 1)$, since r was the minimal natural number satisfying $N \mid (a^r - 1)$. Now, suppose $N \nmid (a^{r/2} + 1)$. Then note that atleast one integer among $\gcd(a^{r/2} - 1, N), \gcd(a^{r/2} + 1, N)$ must be a non-trivial factor of N, and thus we would have succeeded in factorizing N. Also note that $\gcd(a^{r/2} \pm 1, N)$ can be calculated, via Euclid's algorithm, in $\mathrm{poly}(n)$ time. $a^{r/2}$ itself can be calculated in $\mathrm{poly}\log(r) = \mathrm{poly}\log(N)$ time by repeated squaring.
Now, the above reduction makes 3 assumptions: We can find r efficiently [17], r is even, and $N \nmid (a^{r/2}+1)$. Thankfully, over the random choice of a in $[2, N)$, all these properties are satisfied with positive probability. More formally, let a be an uniformly random integer in $[2, N)$, where N is an integer with atleast two prime factors. [18] Conditioned on $\gcd(a, N) = 1$, with probability $\geq 3/8$, the following properties hold:

1. $r := \mathrm{ord}_{\mathbb{Z}_N^\times}(a)$ is even. Note that the minimal natural number r satisfying $a^r \equiv 1 \bmod N$ is just the order of a in the group $\mathbb{Z}_N^\times$.
2. $N \nmid (a^{r/2} + 1)$.

Thus, in expected constant number of trials of the above procedure, we can find an $a \in \mathbb{N}$ such that $\mathrm{ord}_{\mathbb{Z}_N^\times}(a)$ satisfies all the above properties, and thus leads to a factorization of N.
Thus, given $a \in \mathbb{N}$, if we have an efficient way of finding $\mathrm{ord}_{\mathbb{Z}_N^\times}(a)$, we also have an efficient way of factorizing N. Shor's algorithm gives us an efficient way (i.e. $\mathrm{poly}(n)$ time) of finding $\mathrm{ord}_{\mathbb{Z}_N^\times}(a)$.

3.4.1. Period Finding

A function $f : \mathbb{N} \to \mathbb{N}$ is called periodic with period s if $f(x) = f(y)$ if and only if $s \mid (y - x)$, for some $s \in \mathbb{N}$, for all $x, y \in \mathbb{N}$.
Now, let $f : \mathbb{N} \to \mathbb{N}$ be the function $\ell \mapsto (a^\ell \bmod N)$. Note that the period of f is $\mathrm{ord}_{\mathbb{Z}_N^\times}(a)$. Thus, if we have an

[13] By [AKS04], we can determine in $\mathrm{poly}(n)$ time if N is prime. If it is not, then we proceed onto our factorization routine
[14] The best known classical algorithm for factorization runs in $2^{n^{O(1)}}$ time
[15] and if N indeed is a perfect power, then [Ber98] also decomposes N as x^k for $k > 1$. Note that x is a non-trivial factor of N, and we're done
[16] note that $\gcd(a, N)$ can be computed in $\mathrm{poly}(n)$ time by Euclid's algorithm
[17] we need r to calculate $a^{r/2} \pm 1$
[18] N has atleast two prime factors since N is not a perfect power

efficient quantum algorithm for finding the periods of functions, we're done.
The first step of our period finding algorithm is to create a superposition

$$|\psi\rangle := \frac{1}{\sqrt{Q}} \sum_{r=0}^{Q-1} |r\rangle|f(r)\rangle$$

where $Q = 2^q > N^2$ is the smallest power of two strictly greater than N^2. Note that $|r\rangle$ is the binary representation of the integer r written in q qubits. Note that the above superposition can be easily obtained by applying $\mathcal{O}_f \cdot (H^{\otimes q} \otimes I^{\otimes q})$ to $|0\rangle^{2q}$. Now, suppose we measure the last q qubits of $|\psi\rangle$ and obtain $z = f(r)$. Then the first q qubits are in the superposition

$$|\phi_r\rangle := \frac{|r\rangle + |r+s\rangle + \cdots + |r+(L-1)s\rangle}{\sqrt{L}}$$

where s is the period of our function, and $L := \lceil Q/s \rceil$. Also note that when we write $|r+ts\rangle$, we actually mean $|(r+ts) \bmod Q\rangle$.
Thus, given the above state, we have to somehow find s. We do that using the **Quantum Fourier Transform**. Given any integer Q, the $Q \times Q$ quantum Fourier transform is the matrix $F_Q \in \mathbb{C}^{Q \times Q}$ given by

$$(F_Q)_{i,j} := \langle i|F_Q|j\rangle = \frac{\omega^{ij}}{\sqrt{Q}}$$

where $\omega := e^{2\pi\sqrt{-1}/Q}$ is the Q^{th} root of unity. F_Q is clearly unitary; we shall return to the issue of how to implement F_Q later.
Applying F_Q to $|\phi_r\rangle$ yields the state

$$|\tau\rangle := \frac{1}{\sqrt{QL}} \sum_{k=0}^{Q-1} \sum_{\ell=0}^{L-1} \omega^{(r+\ell s)k} |k\rangle = \frac{1}{\sqrt{QL}} \sum_{k=0}^{Q-1} \omega^{rk} \left(\sum_{\ell=0}^{L-1} \omega^{\ell sk} \right) |k\rangle = \frac{1}{\sqrt{QL}} \sum_{k=0}^{Q-1} \omega^{rk} \cdot \left(\frac{1-\omega^{Lsk}}{1-\omega^{sk}} \cdot \mathbb{1}_{\omega^{sk} \neq 1} + L \cdot \mathbb{1}_{\omega^{sk}=1} \right) |k\rangle$$

Before we describe in general what happens, let's look at a special case, when $s \mid Q$: Then note that $L = Q/s$. Now, if $Q \nmid ks$, then $\omega^{ks} \neq 1$, and $\omega^{Lsk} = \omega^{Qk} = 1$. Thus, if $s \mid Q$, then the probability of getting $|k\rangle$ on measuring $|\tau\rangle$, where $Q \nmid ks \iff L \nmid k$, is **zero**. Thus, when we measure $|\tau\rangle$, we obtain one of the outcomes $\{|0\rangle, |L\rangle, \ldots, |(s-1)L\rangle\}$ with equal probability. Thus, by running Shor's algorithm repeatedly, we will obtain outputs of the sort $|x_1 L\rangle, \ldots, |x_t L\rangle$, where $x_1, \ldots, x_t$ are i.i.d samples from $\{0, \ldots, s-1\}$. By suitably adjusting t, with very high probability $\gcd(x_1 L, \ldots, x_t L) = L$. Once we obtain L, we obtain $s := Q/L$, as desired.
Now, in general we wouldn't know if $s \mid Q$ or not. In particular, if $s \nmid Q$, then we wouldn't have perfect interference as above, where the only possible outputs are multiples of L. However, our outputs will still be concentrated around multiples of $L = \lceil Q/s \rceil$.
Thus, write

$$|\tau\rangle = \sum_{k=0}^{Q-1} \alpha_k |k\rangle$$

where $\alpha_k := \frac{1}{\sqrt{QL}} \omega^{rk} \left(\sum_{\ell=0}^{L-1} \omega^{\ell sk} \right)$. We now make precise the notion that if $sk \bmod Q$ is small (i.e. sk is close to a multiple of L), then $|\alpha_k|$ is large. For the following lemmata, assume that $x \bmod Q \in [-Q/2, Q/2] \cap \mathbb{Z}$ for any $x \in \mathbb{Z}$.

Lemma 3.1. If $-s/2 \leq sk \bmod Q \leq s/2$, then $|\alpha_k| \geq \frac{1}{\sqrt{8s}}$.

Proof. Write $\beta := \omega^{sk}$. Note that $|\alpha_k| = \frac{1}{\sqrt{QL}} \left| \sum_{\ell=0}^{L-1} \beta^\ell \right|$, and thus WLOG we focus on bounding the norm of $v := \sum_{\ell=0}^{L-1} \beta^\ell$. Now, note that the angle between the vectors β^0 and β^{L-1} is

$$(L-1) \cdot \frac{2\pi |sk \bmod Q|}{Q} \leq \frac{\pi s(L-1)}{Q} \leq \pi$$

Now, since the angle between β^0 and β^{L-1} is $\leq \pi$, and since v lies along the bisector of the angle created by β^0 and β^{L-1}, atleast half of the terms in the series $\beta^0, \ldots, \beta^{L-1}$ make an angle of $\leq \pi/4$ with v. Since they make an angle of $\leq \pi/4$ with v, they contribute $\geq \cos(\pi/4) = 2^{-1/2}$ to the magnitude of v. Thus

$$|v| \geq \frac{L}{2} \cdot \frac{1}{\sqrt{2}} \implies |\alpha_k| \geq \frac{1}{\sqrt{QL}} \cdot \frac{L}{2\sqrt{2}} \geq \frac{1}{\sqrt{8s}}$$

■

Now, we show that $|sk \bmod Q| \leq s/2$ happens fairly often.

Lemma 3.2. Let $|k\rangle$ be the measurement of $|\tau\rangle$. Then $|sk \bmod Q| \leq s/2$ with probability $\geq 1/16$.

Proof. As k varies over $\{0, \ldots, Q-1\}$, atleast $s/2$ values of $|sk \bmod Q|$ lie in $[-s/2, s/2] \cap \mathbb{Z}$. Each of these k's satisfy $|\alpha_k| \geq \frac{1}{\sqrt{8s}}$. Consequently, with probability $\geq \frac{1}{8s} \cdot \frac{s}{2} = \frac{1}{16}$, $|sk \bmod Q| \leq s/2$. ■

Thus, with probability $\geq 1/16$, we sample a k such that

$$|sk \bmod Q| \leq \frac{s}{2} \iff |sk - cQ| \leq \frac{s}{2} \iff \left|\frac{k}{Q} - \frac{c}{s}\right| \leq \frac{1}{2Q}$$

for some $c \in \mathbb{Z}$.
Now, we get to know $\frac{k}{Q}$ when we measure $|\tau\rangle$. We know that it is very close to $\frac{c}{s}$. Now, clearly $s \leq N$. On the other hand, we have chosen Q to be much larger than N. Thus, even though the fraction itself is corrupted, it is very close to another fraction with a small denominator.
We now describe the continued fraction method for finding out low denominator approximations to any real number. Consider the real number 0.25001. We can write it as:

$$0.25001 = \frac{25001}{100000} = \frac{1}{\frac{100000}{25001}} = \frac{1}{3 + \frac{24997}{25001}} = \frac{1}{3 + \frac{1}{\frac{25001}{24997}}} = \frac{1}{3 + \frac{1}{1 + \frac{4}{24997}}}$$

Note that the "residual fraction" at this point is $\frac{4}{24997}$, which is very small. Thus, if we neglect it, we obtain $0.25001 \approx \frac{1}{3+\frac{1}{1}} = \frac{1}{4}$, i.e. we recover the "uncorrupted" version of 0.25001. Also note that we can calculate the continued fraction expansion of $\frac{k}{Q}$ simply by long division.
We now present a lemma about continued fractions which seals the deal.

Lemma 3.3. Let $\alpha \in \mathbb{R}$ be a real number, and let

$$\alpha_m = a_0 + \frac{1}{a_1 + \frac{1}{\ldots + \frac{1}{a_m}}} = \frac{P_m}{Q_m}, \ \gcd(P_m, Q_m) = 1$$

be the m^{th} continued fraction expansion of α. Then α_m is the best rational approximation of α with denominator $\leq Q_m$, i.e. $|\alpha - \alpha_m| \leq |\alpha - P'_m/Q'_m|$ for all $Q'_m \leq Q_m$, $\gcd(P'_m, Q'_m) = 1$. Furthermore, if $\alpha \in \mathbb{Q}$, then there exists an m such that $\alpha = \alpha_m$.

Thus, given k/Q, we compute the continued fraction expansion of it until the denominator exceeds N. The rational approximation we get thus is c/s: Why? Suppose there were two fractions, c/s and c'/s', such that $\left|\frac{k}{Q} - \frac{c}{s}\right|, \left|\frac{k}{Q} - \frac{c'}{s'}\right| \leq \frac{1}{2Q}$. Then

$$\left|\frac{c}{s} - \frac{c'}{s'}\right| \leq \frac{1}{Q} \implies Q|cs' - c's| \leq ss' \leq N^2$$

However, $|cs' - c's| \geq 1$, and thus $Q \leq N^2$, which is a contradiction.
Thus, the continued fraction approximation of k/Q uniquely recovers c/s. Now, if $\gcd(c, s)$ was 1, then we could recover s too. We argue that this happens with $\frac{1}{\text{poly}(n)}$ probability:

Lemma 3.4. $\gcd(c, s) = 1$ with $\Omega\left(\frac{1}{n}\right)$ probability.

Proof. A little thought reveals that c is uniformly distributed in $\{0, \ldots, s-1\}$. Now, s has at most $\lg s$ prime factors. On the other hand, by the prime number theorem, there are $\Omega(s/\ln s)$ prime numbers in $\{0, \ldots, s-1\}$. Thus there are $\Omega(s/\ln s)$ prime numbers in $\{0, \ldots, s-1\}$ which don't divide s. The probability that c equals one of those primes is $\Omega(1/\ln s) = \Omega(1/n)$, since $s \leq N \implies \ln s = \mathcal{O}(n)$. ■

Thus, with probability $\Omega(1/n)$, we obtain the 'correct' s by measuring and processing the qubits of $(F_Q \otimes I^{\otimes q}) \cdot \mathcal{O}_f \cdot (H^{\otimes q} \otimes I^{\otimes q}) \cdot |0\rangle^{\otimes 2q}$. Also, note that whatever answer we get for s, we can quickly verify if $a^s \equiv 1 \bmod N$. Also, if $a^x \equiv 1 \bmod N$, then $s \mid x$. Thus, suppose we run the period-finding routine T times, and retain only those outputs which satisfy $a^x \equiv 1 \bmod N$. Among those 'correct' x, we choose the minimal value as our s. Then with probability $\Omega(1/n)$, this value for s is indeed the correct value.
Thus, if we run the period-finding routine $T = \Theta(n^2)$ times, then we fail with probability $\leq (1 - \Omega(1/n))^{\Theta(n^2)} = \exp(-\Omega(n))$, i.e. with very high probability we obtain the correct period. Thus, to summarize Shor's algorithm:

Algorithm 1: Shor's Algorithm

Data: N, N is an odd composite number which is not a perfect power
Result: A non-trivial factor of N

1 $a \xleftarrow{\$} [2, N) \cap \mathbb{Z}$
2 **if** $\gcd(a, N) > 1$ **then**
3 **return** $\gcd(a, N)$
4 $s \leftarrow \text{order}(a, N)$ (see Algorithm 2)
5 **if** $2 \nmid s$ **then**
6 **go to** line 1
7 **if** $\gcd(a^{s/2} - 1, N) > 1$ **then**
8 **return** $\gcd(a^{s/2} - 1, N)$
9 **return** $\gcd(a^{s/2} + 1, N)$

Algorithm 2: Period Finding Algorithm

Data: $a, N, \gcd(a, N) = 1$
Result: $\text{ord}_{\mathbb{Z}_N^\times}(a)$

1 Let $Q = 2^q$ be the smallest power of two greater than N^2
2 retval $\leftarrow N$
3 counter $\leftarrow 5\log^2(N)$
4 $|\psi\rangle \leftarrow (F_Q \otimes I^{\otimes q}) \cdot \mathcal{O}_f \cdot (H^{\otimes q} \otimes I^{\otimes q}) \cdot |0\rangle^{\otimes 2q}$
5 Measure first q qubits of $|\psi\rangle$ to obtain $|k\rangle$
6 Let c/s be the continued fraction approximation of k/Q
7 **if** counter > 0 and $a^s \equiv 1 \bmod N$ **then**
8 retval $\leftarrow \min(\text{retval}, s)$
9 counter $\leftarrow$ counter -1
10 **go to** line 4
11 **return** retval

3.4.2. Implementation Issues

We'll now see how to implement F_Q and $\mathcal{O}_f$. Implementing $\mathcal{O}_f$ is easy: Note that $f(\ell) := a^\ell \bmod N$ is a function that can be computed easily classically in $\operatorname{poly}\log(\ell)$ steps. Since the maximum ℓ we invoke $\mathcal{O}_f$ for is $Q-1$, we can implement $\mathcal{O}_f$ in $\operatorname{poly}\log(Q) = \operatorname{poly}(n)$ gates. The classical circuit to compute f also works as the oracle $\mathcal{O}_f$ (which simply treats the qubits as bits).
Implementing F_Q is much trickier: Firstly, we note that

$$F_Q = \frac{1}{\sqrt{2}}\begin{bmatrix} F_{Q/2} & B_{Q/2}F_{Q/2} \\ F_{Q/2} & -B_{Q/2}F_{Q/2}\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} I_{Q/2} & I_{Q/2} \\ I_{Q/2} & -I_{Q/2}\end{bmatrix}\cdot\begin{bmatrix} I_{Q/2} & 0 \\ 0 & B_{Q/2}\end{bmatrix}\cdot\begin{bmatrix} F_{Q/2} & 0 \\ 0 & F_{Q/2}\end{bmatrix} = (H\otimes I_{Q/2})\cdot\begin{bmatrix} I_{Q/2} & 0 \\ 0 & B_{Q/2}\end{bmatrix}\cdot(I_2\otimes F_{Q/2})$$

where

$$B_{Q/2} := \begin{bmatrix} 1 & & & & \\ & \omega & & & \\ & & \omega^2 & & \\ & & & \ddots & \\ & & & & \omega^{Q/2-1}\end{bmatrix}$$

is a diagonal matrix. Note that although the above matrix relations for F_Q might appear mysterious, they're in fact quite straightforward: The first equality simply expresses a recursion, in a condensed, matrix form. The other equalities are then simply matrix manipulations.
It is easy to convince oneself that if we can implement $B_{Q/2}$, then we can implement F_Q too. Implementing $B_{Q/2}$ is easy: If the first qubit is 1, then some rotation is applied. If the second qubit is 1, some other rotation is applied, and so on. Thus, we only require $\operatorname{poly}\log Q = \operatorname{poly}(n)$ gates to implement $B_{Q/2}$.
With some effort, one can show that F_Q itself can be implemented in $\operatorname{poly}\log Q = \operatorname{poly}(n)$ gates, as desired.

3.5. Hidden Subgroup Problem

Shor's algorithm can be interpreted as trying to find a "hidden" subgroup in an abelian group. Before we elucidate on this connection further, let's define the Hidden Subgroup problem:

Problem (Hidden Subgroup Problem (HSP))**.** Let G be a group, and let H be a subgroup of G. We are given a function $f : G \to S$, and we're promised that f assigns a unique element to each coset of H, i.e. f restricts to an injective function $G/H \to S$. Our task is to describe H (by producing a generating set for H), given oracle access to f.

Remark. Note that if $|H| = \ell$, then there exists a generating set for H of size $\leq \lg \ell$: Indeed, define a set of groups inductively as $H_0 := \{\mathrm{id}_G\}$, $H_i := \langle H_{i-1}, x_{i-1}\rangle$, for $i \geq 1$, where $x_{i-1} \in H \setminus \langle H_{i-1}\rangle$. Since x_{i-1} forms atleast two distinct cosets of H_{i-1}, $|H_i| \geq 2|H_{i-1}|$, and we're done.
All HSPs we'll see have group sizes bounded by $\exp(\operatorname{poly}(n))$ for relevant parameter n, and thus a $\operatorname{poly}(n)$ generating set will always exist.

This is how we reduce Simon's and Shor's problems to HSPs:

Problem (Simon's Problem)**.** Let $G = \mathbb{Z}_2^n$. G has a subgroup $H := \{0, s\}$. We are given a function $f : G \to G$ which assigns a unique element to every coset of H. Our task is to find H.

It is easy to see the equivalence of this formulation with the one we saw earlier.

Problem (Shor's Problem)**.** Let $G = \mathbb{Z}_N$. Let $a \in \mathbb{Z}_N^\times$ be some given integer, and let $r = \mathrm{ord}_{\mathbb{Z}_N^\times}(a)$. Let $H = \langle r \rangle$ be the subgroup generated by the element $r \in G$. Let $f : G \to G$ be the function given by $g \to a^g \bmod N$.

Turns out we can encode even more problems using this:

Problem (Discrete Logarithm)**.** Let $G = \mathbb{Z}_n \times \mathbb{Z}_n$. Let $G' = \langle g \rangle$ be another group such that $|G'| = n$, and let $h \in G'$. Let $H := \{(x, y) \in \mathbb{Z}_n \times \mathbb{Z}_n : g^x h^y = 1\}$ be a subgroup of G. Consider the function $f : G \to G'$ given by $f(a, b) := g^a h^b$.

One can also show that the Graph Isomorphism problem [19] reduces to a HSP over $\mathfrak{S}_n$. One can also show that the Shortest Vector Problem [20] reduces to a HSP over the dihedral group D_n.
We shall now show how quantum computing can be used to give efficient solutions for all HSPs on *abelian* groups. Many of these abelian HSPs have no known efficient classical algorithms. Despite intense research, efficient quantum algorithms haven't been found for HSPs over *non-abelian groups*. In particular, the Graph Isomorphism Problem, or the Shortest Vector Problem, remain out of reach for polynomial time quantum algorithms.
Let $N = 2^n$ be the smallest power of two greater than or equal to $|G|$. We embed G into $\{0, 1\}^n$. Let $f : G \to S$ be our HSP function. Note that WLOG $|S| = |G/H| \leq |G| \leq 2^n$. Thus we can embed S into $\{0, 1\}^n$ as well. For $g \in G \subseteq \{0, 1\}^n$, and $x \in \{0, 1\}^n$, define the oracle

$$\mathcal{O}_f(|g\rangle \otimes |x\rangle) := |g\rangle \otimes |x \oplus f(g)\rangle$$

For $y \in \{0, 1\}^n \setminus G$, we leave $\mathcal{O}_f(|y\rangle \otimes |x\rangle)$ undefined.
While we shall not go into the details of our algorithm, note that the key ingredient in Shor's algorithm was the Quantum Fourier Transform (QFT). Using ideas similar to that, we can implement Quantum Fourier Transform on all abelian groups. HSPs for abelian groups then follow by massaging the Fourier transform into desired states.

3.6. Grover's Algorithm [Gro96]

We have the following problem: We are given N elements $X := \{x_1, \ldots, x_N\}$ in our database, and a function $f : X \to \{0, 1\}$. We are promised that $f^{-1}(1)$ is non-empty. We call the elements of $f^{-1}(1)$ "marked items". We have to find an element $x \in X$ such that $f(x) = 1$.
Right off the bat, we make a few assumptions: We assume $N = 2^n$ is a power of two. This can be ensured easily by adding garbage elements to X if necessary, and setting their f-values to 0. Since $N = 2^n$, we may also assume that $X = \{0, 1\}^n$. Finally, we assume that there is a unique $x_* \in X$ such that $f(x_*) = 1$. [21]
Note that any classical algorithm (even randomized) for this must necessarily take $\Omega(N)$ time. We now provide a quantum algorithm, called Grover's algorithm, which achieves this in $\mathcal{O}(\sqrt{N})$ time.
Our input is $|0\rangle^{\otimes n}$. The initial steps of the algorithm are quite obvious: Apply $H^{\otimes n}$ to $|0\rangle^{\otimes n}$ to obtain

$$|\psi\rangle := \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \{0,1\}^n} |x\rangle$$

At this point, we introduce the so-called *Grover Diffusion Operator*: If D is the diffusion operator, and if $|\alpha\rangle := \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle$ is a state, then

$$D\alpha := \sum_{x \in \{0,1\}^n} (2\overline{\alpha} - \alpha_x)|x\rangle$$

[19] which gives two isomorphic labeled graphs and asks us to find a bijection between the labels which takes one graph to the other
[20] which gives us a *lattice* $L := \{a_0 z_0 + \cdots + a_{n-1} z_{n-1} : a_0, \ldots, a_{n-1} \in \mathbb{Z}\}$, where $z_0, \ldots, z_{n-1}$ are given vectors in $\mathbb{R}^n$. Our task is to find a $\mathcal{O}(\sqrt{n})$-approximation to the shortest vector in L
[21] this assumption can be removed without much difficulty

where $\overline{\alpha} := 2^{-n} \sum_{x \in \{0,1\}^n} \alpha_x$ is the average of the coefficients.
In other words, D flips all coefficients about the mean. The explicit form of D is

$$D := \begin{bmatrix} \frac{2}{N} - 1 & \frac{2}{N} & \cdots & \frac{2}{N} \\ \frac{2}{N} & \frac{2}{N} - 1 & \cdots & \frac{2}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{2}{N} & \frac{2}{N} & \cdots & \frac{2}{N} - 1 \end{bmatrix}$$

It can be easily verified that D is unitary. We shall return to the issue of actually implementing D later.
Then Grover's Algorithm is as follows: Apply $(D \cdot \mathcal{O}_f)^T$ to $|\psi\rangle$, for some carefully chosen parameter T.
Informally speaking, $\mathcal{O}_f$ first flips the amplitude for $|x_*\rangle$, while keeping all others as it is. Since only the amplitude for $|x_*\rangle$ gets flipped, the mean of the amplitudes stays $\approx \frac{1}{\sqrt{N}}$. Now, when D flips all coefficients about the mean, since the coefficients of $|x\rangle, x \neq x_*$ are approximately equal to the mean anyways, they stay put. However, the amplitude of $|x_*\rangle$ gets a boost, since it was of the opposite sign as the mean. Repeating this procedure (of applying $\mathcal{O}_f$ and then D) many times boosts the amplitude of $|x_*\rangle$ to $\Omega(1)$, while the other amplitudes stay at $\mathcal{O}(1/\sqrt{N})$. At this point, measuring our state outputs x_* with good probability.
To formally analyze Grover's algorithm, denote by $\alpha^{(t)}$ the amplitude of $|x_*\rangle$ *after* t steps, and let $\beta^{(t)}$ be the amplitudes of the other elements. Then $\alpha^{(0)} = \beta^{(0)} = \frac{1}{\sqrt{N}}$. Also note that $\alpha^{(t+1)} = -\alpha^{(t)} + \frac{2(N-1)}{N}\beta^{(t)}$: Indeed, after application of $\mathcal{O}_f$, the amplitude of $|x_*\rangle$ becomes $-\alpha^{(t)}$. The mean of the amplitudes at this point is $\frac{-\alpha^{(t)} + (N-1)\beta^{(t)}}{N}$. Thus, at the next step, we have

$$\alpha^{(t+1)} = 2 \cdot \frac{-\alpha^{(t)} + (N-1)\beta^{(t)}}{N} + \alpha^{(t)} = \frac{N-2}{N}\alpha^{(t)} + \frac{2(N-1)}{N}\beta^{(t)}$$

Now, $(\alpha^{(t)})^2 + (N-1)(\beta^{(t)})^2 = 1$. Thus $\beta^{(t)} = \pm\sqrt{\frac{1-(\alpha^{(t)})^2}{N-1}}$. Assume $\beta^{(t)} \geq 0$. Then

$$\alpha^{(t+1)} = \frac{N-2}{N}\alpha^{(t)} + \frac{2\sqrt{N-1}}{N}\sqrt{1-(\alpha^{(t)})^2}$$

Write $\alpha^{(t)} := \sin(\theta^{(t)})$, $\theta := \cos^{-1}\left(\frac{N-2}{N}\right)$. Then

$$\sin(\theta^{(t+1)}) = \sin(\theta^{(t)} + \theta) \implies \theta^{(t+1)} = \theta^{(t)} + \theta = \theta^{(0)} + (t+1)\theta$$

Thus, for small t (to ensure $\beta^{(t)}$ stays positive), we have $\theta^{(t)} = \sin^{-1}\left(\frac{1}{\sqrt{N}}\right) + t \cdot \cos^{-1}\left(\frac{N-2}{N}\right)$. Thus, if we set

$$T := \left\lfloor \frac{\cos^{-1}\left(\frac{1}{\sqrt{N}}\right)}{\cos^{-1}\left(\frac{N-2}{N}\right)} \right\rfloor \sim \frac{\pi}{4} \cdot \sqrt{N}$$

After application of T steps, $\frac{\pi}{2} - \theta \leq \theta^{(T)} \leq \frac{\pi}{2}$, and thus $\alpha^{(t)} \geq \sin(\pi/2 - \theta) = \cos(\theta) = 1 - 2/N$. Consequently, after applying $(D \cdot \mathcal{O}_f)^T$, if we measure the resulting state, we will obtain x_* with probability $1 - \mathcal{O}(1/N)$.
Some interesting points about Grover's algorithm:

1. Suppose we let Grover's algorithm run too long, say upto $T = \frac{\pi}{2}\sqrt{N}$. Then the probability of observing x_* goes to 0! Thus Grover's algorithm is an example of a randomized algorithm whose success probability drops if we let it run too long!

2. Suppose $|f^{-1}(1)| = K$. Then the optimal stopping time T to observe an element of $f^{-1}(1)$ is $\frac{\pi}{4}\sqrt{\frac{N}{K}}$. So what is the optimal stopping time when K is unknown? We binary search over the possible values of K: First assume K is N, run Grover's algorithm for $\pi/4\sqrt{N/K}$ steps. If we find a marked item, we stop. Otherwise assume K is $N/2$, and run Grover's algorithm for $\pi/4\sqrt{N/K}$ steps. We continue this process until we go down to $K = 1$. This process still takes $\mathcal{O}(\sqrt{N/K_*})$ queries, where $K_* = |f^{-1}(1)|$ is the number of marked items (which we don't know, and don't need to know).

3. Note that if multiple runs of Grover's algorithm (suitably adapted for various K) fail to find any marked item, then with high probability we can be certain that there is no marked item. More precisely, since Grover search succeeds in finding a marked item (if there is any) with $\Omega(1)$ probability, by running Grover's algorithm $\mathcal{O}(1)$ many times, we can also certify with $\Omega(1)$ probability if there is any marked item.

3.6.1. Implementation Issues

How do we actually implement D? Note that $D = H^{\otimes n} A H^{\otimes n}$, where $A = \begin{bmatrix} 1 & & & \\ & -1 & & \\ & & \ddots & \\ & & & -1 \end{bmatrix} = \langle 0|^n |0\rangle^n - I$. Thus, if we can implement A, we're done. Note that for any $x \in \{0,1\}^n$,

$$A|x\rangle = \begin{cases} -|x\rangle & \text{if } x = 0^n \\ |x\rangle & \text{otherwise} \end{cases}$$

Thus, A negates $|x\rangle$ if and only if $x = 0^n$. Note that it is easy to implement A as a classical circuit: We first apply the n-bit OR gate on the input, and if the output is 0, we apply $-I$, else we apply I.
We can implement the above circuit using $\mathcal{O}(n) = \mathcal{O}(\lg N)$ Toffoli gates. Since $H^{\otimes n}$ can also be implemented using $\mathcal{O}(\lg N)$ gates, the Grover diffusion operator takes $\mathcal{O}(\lg N)$ gates to implement, and Grover's algorithm itself takes $\mathcal{O}(\sqrt{N} \lg N)$ gates to implement.

3.7. Applications of Grover's Algorithm

3.7.1. OR-of-ANDs

Suppose we have N input bits arranged in a $\sqrt{N} \times \sqrt{N}$ table. We have to find out if there is any row in the table which contains only ones. This type of a problem is known as an OR-of-ANDs problem: Indeed, if the bits of our table are arranged as $\{x_{ij}\}$, then the above problem asks if the following OR-of-ANDs expression is 1:

$$\bigvee_{i=1}^{\sqrt{N}} \bigwedge_{j=1}^{\sqrt{N}} x_{ij}$$

Once again, any classical algorithm for this takes $\Omega(N)$ time. Grover allows us to do this in $\widetilde{\mathcal{O}}(\sqrt{N})$ time: In $\mathcal{O}(\sqrt{\sqrt{N}}) = \mathcal{O}(N^{1/4})$ time, we can search if any given row has a 0. We call a row 'marked' if running Grover on that row failed to find a 0. We now run an 'outer' Grover over all rows, trying to find a marked row. The oracle for the outer Grover checks if a row is marked by running the above 'inner' Grover on that row in $\mathcal{O}(N^{1/4})$ time.
Thus, in $\mathcal{O}(N^{1/4} \cdot N^{1/4}) = \mathcal{O}(N^{1/2})$ time, we can find if a row contains only ones. We run the Grover routines polylog many times to reduce error to $1/\operatorname{poly}(N)$, thus taking our run-time to $\widetilde{\mathcal{O}}(\sqrt{N})$. With some cleverness, these poly-log factors can be removed, and thus the OR-of-ANDs problem can be solved in $\mathcal{O}(\sqrt{N})$ time.
In fact, suppose we have an OR-of-ANDs tree, i.e. a tree whose root is an OR, and the rest of the tree is an alternating pattern of ORs and ANDs, with the variables at the leaves. Also assume there are N leaves. Then a result due to Farhi, Goldstone and Gutmann shows that one can solve the OR-of-ANDs problem in $\mathcal{O}(\sqrt{N})$ time. However, the general case is much trickier than the one above, since controlling errors in Grover's algorithm across arbitrary depths becomes much tougher.

3.7.2. The Collision Problem

Let N be an even integer, and let $f : [N] \to [N]$ be a 2-to-1 function, i.e. every element in the image of f has exactly two pre-images. Our task is to find $x, y, x \neq y$ such that $f(x) = f(y)$.
A typical birthday paradox argument gives us a $\mathcal{O}(\sqrt{N})$ algorithm for the above problem, and the adversary method also yields a $\Omega(\sqrt{N})$ lower bound for classical problems.

On the other hand, the naive quantum algorithm also runs in $\mathcal{O}(\sqrt{N})$ query complexity: Indeed, run Grover's algorithm on $[N-1]$ to find an element x such that $f(x) = f(N)$.
So can we do better? Once again, a "layered" Grover gets us what we want: Randomly choose $N^{1/3}$ elements from $[N]$, and query the oracle for their values. Once we have their f-values, sort them to enable fast lookup. Let $\mathcal{F}$ be the sorted set of these f-values. Note that if we detect a collision within $\mathcal{F}$, then we're already done.
Now, choose $N^{2/3}$ elements from $[N] \setminus \{\text{the elements already chosen earlier}\}$. An element x chosen in this round is marked if $f(x) \in \mathcal{F}$. Run Grover on these $N^{2/3}$ elements to find a marked element. Now, since the above $N^{1/3} + N^{2/3}$ elements were chosen randomly, with $\Omega(1)$ probability there is a marked element among the $N^{2/3}$ elements. [22] Furthermore, Grover will find that element in $\mathcal{O}(\sqrt{N^{2/3}}) = \mathcal{O}(N^{1/3})$ queries. Thus, the total number of queries is $N^{1/3} + \mathcal{O}(N^{1/3}) = \mathcal{O}(N^{1/3})$, i.e. with $\Omega(1)$ probability we can find a collision in $\mathcal{O}(N^{1/3})$ queries.

3.7.3. Element Distinctness

Here, we're given a function $f : [N] \to [N]$. We have to determine if f is injective.
Classically, we can solve this problem in $\widetilde{\mathcal{O}}(N)$ time by hashing/sorting.
Quantumly, we can solve this in $\widetilde{\mathcal{O}}(N^{3/4})$ time using a "layered" Grover as above. However, as we shall see later, the optimal algorithm for this problem comes via *quantum random walks*, and takes $\mathcal{O}(N^{2/3})$ time.

[22] indeed, with high probability $|\mathcal{F}| = \Omega(N^{1/3})$. We call $x \neq y$ partners if $f(x) = f(y)$. Note that none of the partners of the $\Omega(N^{1/3})$ elements chosen in the first round were sampled in the $N^{1/3} + N^{2/3}$ samples. The probability of this occurring is $\leq (1 - \Omega(N^{1/3})/N)^{N^{1/3}+N^{2/3}} = \mathcal{O}(1)$

§4. Lower Bounds against Quantum Algorithms

We have seen many quantum algorithms; we'll now see matching lower bounds for many of them. Before we do that, we define some terminology:

1. **Search Problems**: We call a problem a search problem if the output is some string/object.
2. **Decision Problems**: We call a problem a decision problem if the output is a yes/no.
3. **Promise Problems**: We call a problem a promise problem if the class of possible input functions is promised to have some property.
4. **Total Problems**: Problems which are not promise problems are called total problems.

A very exciting theorem says that for total decision problems, quantum algorithms can't offer an exponential speedup over classical algorithms!

Theorem 4.1 ([ABDK+21, ABB+17, BS21, SSW23])**.** Let P be a total decision problem with parameter N. Let $D(N), R(N), Q(N)$ be the deterministic, randomized, and quantum query complexity of P respectively. Then $D(N) = \mathcal{O}(Q(N)^4)$, which automatically implies $R(N) = \mathcal{O}(Q(N)^4)$.
Furthermore, there exists a promise problem for which $D(N) = \Omega(Q(N)^4)$. There also exists a promise problem for which $R(N) = \Omega(Q(N)^3)$.

Note that both Grover's problem and the Element Distinctness (ED) problem can be viewed as total decision problems: For Grover, we have to decide if a database contains any marked item, and for the Element Distinctness problem we have to decide if the function is injective. In both cases, there are no promises on the database or the function. Thus, the above theorem explains why we don't get exponential speedups for Grover and ED like we do for Simon's problem.
We shall now see the polynomial method for proving lower bounds.

4.1. The Polynomial Method

Instead of assuming our input is a function $f : [N] \to [M]$, we assume its a string $w \in [M]^N$, which we have oracle access to. Our oracle to w will be denoted $\mathcal{O}_w$, and depending on the context, $\mathcal{O}_w(|i\rangle) = (-1)^{w_i}|i\rangle$ (if $[M] \cong \{0,1\}$), or $\mathcal{O}_w(|i\rangle|y\rangle) = |i\rangle|y \oplus f(i)\rangle$ (if $M \cong \{0,1\}^m, N \cong \{0,1\}^n$). Also denote by $\widetilde{w_{i,c}}$, where $i \in [N], c \in [M]$, the indicator variable:

$$\widetilde{w_{i,c}} := \begin{cases} 1 & \text{if } w_i = c \\ 0 & \text{otherwise} \end{cases}$$

We now prove a basic statement about the polynomial method.

Lemma 4.2. Let $\mathcal{A}$ be a quantum query algorithm which makes t queries to $\mathcal{O}_w$, where $w \in [M]^N$. Also assume $[N] \cong \{0,1\}^n, [M] \cong \{0,1\}^m$. Then the amplitude of any basis state after the application of $\mathcal{A}$ is a multilinear polynomial in $\widetilde{w} := \{\widetilde{w_{j,c}}\}_{j\in[N],c\in[M]}$ of degree $\leq t$.

Proof. WLOG the input to $\mathcal{A}$ is $|0\rangle^{\otimes n} \otimes |0\rangle^{\otimes m} \otimes |0\rangle^{\otimes a}$, where n is the number of input qubits/registers, m is the number of output registers, and a is the number of ancilla qubits. Then note that

$$\mathcal{A} = \prod_{i=0}^{t} (\mathcal{O}_w \cdot U_i)$$

where $U_0, \ldots, U_t$ are some unitaries.
WLOG it suffices to carry out all measurements at the end of $\mathcal{A}$ by the Deferred Measurement Principle (see Section 1.12).
After application of U_0, the state becomes

$$\sum_{j\in\{0,1\}^n} \sum_{b\in\{0,1\}^m} \sum_{z\in\{0,1\}^a} \alpha_{jbz}|j\rangle|b\rangle|z\rangle$$

Note that α_{jbz} is a constant (i.e. polynomial of degree 0) w.r.t $\widetilde{w}$, because $\mathcal{O}_w$ hasn't been invoked yet. Thus the base case of the induction (for $t = 0$) is verified.
Now, suppose after k steps our state looks like

$$|\psi\rangle = \sum_{j\in\{0,1\}^n} \sum_{b\in\{0,1\}^m} \sum_{z\in\{0,1\}^a} P_{jbz}(\widetilde{w})|j\rangle|b\rangle|z\rangle$$

where $P_{jbz}(\widetilde{w})$ is a polynomial of degree $\leq k$. Now, note that applying a unitary transformation doesn't change the degree w.r.t. $\widetilde{w}$ [23]. Thus, we can focus on what happens when $\mathcal{O}_w$ is applied:

$$\mathcal{O}_w \cdot |\psi\rangle = \sum_{j\in\{0,1\}^n} \sum_{b\in\{0,1\}^m} \sum_{z\in\{0,1\}^a} P_{jbz}(\widetilde{w})|j\rangle|b \oplus w_j\rangle|z\rangle$$

$$= \sum_{j\in\{0,1\}^n} \sum_{b,b'\in\{0,1\}^m} \sum_{z\in\{0,1\}^a} \widetilde{w_{j,b'\oplus b}} \cdot P_{jbz}(\widetilde{w})|j\rangle|b'\rangle|z\rangle = \sum_{j\in\{0,1\}^n} \sum_{b'\in\{0,1\}^m} \sum_{z\in\{0,1\}^a} \underbrace{\left(\sum_{b\in\{0,1\}^m} \widetilde{w_{j,b'\oplus b}} \cdot P_{jbz}(\widetilde{w}) \right)}|j\rangle|b'\rangle|z\rangle$$

It is clear that the underbraced polynomial is of degree $\leq k+1$, as desired.
Finally, to see that we can take $P_{jbz}(\widetilde{w})$ to be multilinear, observe that $\widetilde{w_{j,c}}^2 = \widetilde{w_{j,c}}$, since $\widetilde{w_{j,c}}$ is a $0-1$ valued variable. Consequently, we can multilinearize any polynomials we obtain and assume WLOG that P_{jbz} is multilinear. ■

Corollary 4.3. Let $\mathcal{A}$ be a quantum query algorithm which makes t queries to $\mathcal{O}_w$, where $w \in [M]^N$. Also assume $[N] \cong \{0,1\}^n, [M] \cong \{0,1\}^m$. Then the acceptance probability of $\mathcal{A}$ is a multilinear polynomial in $\widetilde{w}$ of degree $\leq 2t$.

Proof. The probability of obtaining $|j\rangle|b\rangle|z\rangle$ is $P_{jbz} \cdot P^*_{jbz}$. Since $\deg(P_{jbz}) \leq t$, the result follows. ■

Suppose we have a decision problem, i.e. given a string $w \in [M]^N$, we have to decide if w satisfies some property $\mathcal{P}$. Thus the property $\mathcal{P}$ is a function $\mathcal{P} : [M]^N \to \{0,1\}$. Now, let $\mathcal{A}$ be a quantum query algorithm for this problem, and suppose $\mathcal{A}$ succeeds with probability $\geq 2/3$. Let $P(\widetilde{w})$ be the degree $\leq 2t$ polynomial given by Corollary 4.3, i.e. $P(\widetilde{w})$ gives us the probability that $\mathcal{A}$ accepts. Then note that for all $w \in [M]^N$, we have:

$$|P(\widetilde{w}) - \mathcal{P}(w)| \leq \frac{1}{3}$$

4.1.1. Lower Bounds for Grover's Problem

Note that Grover's decision problem can be framed as follows: Given a database $\{w_1, \ldots, w_N\}$, where $w_i \in \{0,1\}$, decide if $\mathrm{OR}(w_1, \ldots, w_N) = 1$. Note that we say w_i is marked if and only if $w_i = 1$. Also note that $M = \{0,1\}$ for this problem.
Let $P(w_1, \ldots, w_N)$ be the accepting polynomial of a quantum algorithm for Grover decision with $\leq t$ queries. Note that technically P should be a polynomial of $\widetilde{w_{j,c}}$. However, since $c \in \{0,1\}$, $\widetilde{w_{j,1-c}} = 1 - \widetilde{w_{j,c}}$. Thus, it is enough to consider P to be a polynomial of $\widetilde{w_{j,1}}$. But note that $\widetilde{w_{j,1}} = w_j$, and thus $P(\widetilde{w}) = P(w_1, \ldots, w_N)$.

[23] Recall that we are using $\widetilde{w}$ as a shorthand for $\{\widetilde{w_{j,c}}\}_{j\in[N],c\in[M]}$

We have $P(0,\dots,0) \in [0,\frac{1}{3}]$ and $P(w_1,\dots,w_N) \in [\frac{2}{3},1]$ if any of $w_1,\dots,w_N$ is 1, by the guarantees of our algorithm. Define the *symmetrization* of P to be:

$$Q(w_1,\dots,w_N) := \frac{1}{N!}\sum_{\pi\in\mathfrak{S}_N} P(w_{\pi(1)},\dots,w_{\pi(N)})$$

Note that $\deg(Q) \leq \deg(P)$. Also note that $Q(0,\dots,0) \in [0,\frac{1}{3}]$, and $Q(w_1,\dots,w_N) \in [\frac{2}{3},1]$ if any of $w_1,\dots,w_N$ is 1. Thus for any $w \in \{0,1\}^N$,

$$\big|Q(w) - \mathrm{OR}(w)\big| \leq \frac{1}{3}$$

where $\mathrm{OR}(w) := \mathrm{OR}(w_1,\dots,w_N) \in \{0,1\} \subseteq \mathbb{R}$ is the standard OR function.
Now, note that $Q(w_1,\dots,w_N)$ is a symmetric function: It doesn't change if we permute its arguments. Consequently, if we just specify the number of ones among $\{w_1,\dots,w_N\}$, we should be able to compute $Q(w)$. In other words Q is actually just a function of the Hamming weight of w. Thus, we can view Q as a univariate polynomial over the integers. We formalize this through the lemma below:

Lemma 4.4. Let $Q : \{0,1\}^N \to \mathbb{R}$ be a symmetric multilinear polynomial of degree d. Then there exists a polynomial $q : \mathbb{R} \to \mathbb{R}$ such that $Q(w) = q(\mathrm{wt}(w))$, and $\deg(q) \leq d$.

Proof Sketch. Since Q is a symmetric multilinear polynomial, it is easy to see by induction that

$$Q = \sum_{k=0}^{d} \alpha_k \sum_{T\in\binom{[N]}{k}} w_T$$

where $w_T := \prod_{i\in T} w_i$, $\alpha_0,\dots,\alpha_d \in \mathbb{R}$. Now, write $p_k := \sum_{i=1}^N w_i^k$. By Newton's identities, $\sum_{T\in\binom{[N]}{k}} w_T = p(p_0,p_1,\dots,p_k)$, where p is a polynomial of degree $\leq k$. On the other hand, since the w_i are $0-1$ valued variables, $p_k = p_1$ for all $k \geq 1$. Thus, restricted to $\{0,1\}^N$, one can express Q as a univariate polynomial of p_1 (note that $p_0 = N$ is just a constant) of degree $\leq d$. But $p_1 = \sum_{i=1}^N w_i = \mathrm{wt}(w)$, and thus we're done. ∎

The polynomial q obtained in the above lemma satisfies the property that $q(0) \in [0,\frac{1}{3}]$, $q(z) \in [\frac{2}{3},1]$ for all $z \in [N]$. If we can show that any polynomial satisfying this property necessarily has a large degree, then we're done. For that, we shall need Markov's "other" inequality:

Theorem 4.5 (Markov's Other Inequality)**.** Let $q : \mathbb{R} \to \mathbb{R}$ be a univariate polynomial of degree d. Let I be an interval of length ℓ, and let $h := \max_{x\in I} |q(x)|$. Then for any $x \in I$,

$$|q'(x)| \leq \frac{d^2 h}{\ell}$$

Now, write $h := \max_{x\in[0,N]} |q(x)|$. We now make cases:

1. $h \leq 2$: Then by Theorem 4.5, for any $x \in [0,N]$, $|q'(x)| \leq \frac{2(\deg(q))^2}{N}$. On the other hand, since $q(0) \leq 1/3, q(1) \geq 2/3$, by the mean value theorem there exists some $y \in [0,1] \subseteq [0,N]$ such that $q'(y) \geq \frac{1}{3}$. Thus
$$\frac{1}{3} \leq \frac{2(\deg(q))^2}{N} \implies \deg(q) \geq \sqrt{\frac{N}{6}}$$
2. $h > 2$: Let $y \in [0,N]$ be such that $|q(y)| = h$. Let y' be the integer in $[0,N]$ closest to y. Then $|q(y')| \leq 1$. Also, $|y - y'| \leq \frac{1}{2}$. Thus, by the mean value theorem, there exists y'' in between y, y' such that $|q(y'')| \geq 2(h-1) > 2$. Thus $2 \leq \frac{2(\deg(q))^2}{N}$, and we have $\deg(q) \geq \sqrt{N}$.

Thus $\deg(q) \geq \sqrt{\frac{N}{6}}$. But also note that $\deg(q) \leq \deg(Q) \leq \deg(P) \leq 2t$, where t is the number of queries our algorithm made. Thus we have the following result:

Theorem 4.6 (Lower Bound for Grover Decision)**.** Let $\{w_1, \ldots, w_N\}$ be a database, and suppose $\mathcal{A}$ is a quantum algorithm, which, with probability $\geq 2/3$, determines if the database contains a 1. Then $\mathcal{A}$ must necessarily make $\geq \frac{\sqrt{N}}{2\sqrt{6}} = \sqrt{\frac{N}{24}}$ queries to $\mathcal{O}_w$.

Now, suppose $\mathcal{A}$ is a quantum algorithm for Grover **search**, instead of Grover **decision**. Note that if $\mathcal{A}$ can find a marked item (if there is any) with probability $\geq 2/3$, then by running $\mathcal{A}$ $\mathcal{O}(1)$ many times, we can decide with probability $\geq 2/3$ if our database has any marked item in the first place. Consequently, the lower bound for Grover decision transfers to Grover search, and we have:

Theorem 4.7 (Lower Bound for Grover Search)**.** Let $\{w_1, \ldots, w_N\}$ be a database, and suppose $\mathcal{A}$ is a quantum algorithm, which, with probability $\geq 2/3$, finds out a i such that $w_i = 1$, if there is any such i, or says that our database doesn't have any marked item. Then $\mathcal{A}$ must necessarily make $\Omega(\sqrt{N})$ queries to $\mathcal{O}_w$.

Consequently, Grover's algorithm is optimal upto constant factors.

4.1.2. Lower Bounds for the Collision Problem and Element Distinctness Problem

Recall the collision problem: Given $f : [N] \to [M]$, we have to decide if f is injective, or if it is 2-to-1, i.e. every element in the image of f has exactly 2 pre-images. Note that this is a promise problem, since we're promised that f is either injective or 2-to-1.
By the algorithm presented in Section 3.7.2, we can detect a collision with positive probability, if there is any, in $\mathcal{O}(N^{1/3})$ time. Consequently, if we don't find a collision even after running our collision-finding algorithm some (large) constant many times, then with high probability we can certify that f is injective, also in $\mathcal{O}(N^{1/3})$ time.
We shall now prove a matching lower bound, i.e. any algorithm deciding if a given $f : [N] \to [M]$ is injective or 2-to-1, must necessarily make $\Omega(N^{1/3})$ queries to $\mathcal{O}_f$ (or $\mathcal{O}_w$).
Before we proceed onto the proof of this, the above lower bound for the collision problem immediately implies a lower bound for the Element Distinctness problem.

Theorem 4.8 (Lower Bound for Element Distinctness)**.** Suppose we are given $f : [N] \to [N]$, and we want to find out if f is injective. Then any quantum algorithm answering the above problem correctly with probability $\geq 2/3$ must necessarily make $\Omega(N^{2/3})$ queries to $\mathcal{O}_f$.

Proof. AFTSOC it was possible to decide the injectivity of f in $o(N^{2/3})$ queries. We derive a contradiction by violating the lower bound for the collision problem.
Now, let $f : [N] \to [M]$ be our input for the collision problem: We're promised that f is either injective or 2-to-1. Let $w \in [M]^N$ be the word form of f. Randomly sample $\sqrt{N}$ elements from $[N]$. With high probability, $\Omega(\sqrt{N})$ of the sampled elements are distinct.
Now, if w is injective, then $w|_S$ is also injective. Conversely, if w is 2-to-1, then with probability $\geq 1-(1-\sqrt{N}/N)^{\sqrt{N}} = \Omega(1)$ $w|_S$ is also not injective. Now, invoke the ED algorithm to decide if $w|_S$ is injective in $o((N^{1/2})^{2/3}) = o(N^{1/3})$ queries. Thus, with $\Omega(1)$ probability we're able to decide if w is injective in $o(N^{1/3})$ queries, leading to a contradiction. ■

Remark. Note that we only saw a $\mathcal{O}(N^{3/4})$ algorithm for the ED problem. However, a $\mathcal{O}(N^{2/3})$ algorithm for the ED problem was given by Ambainis [Amb07], thus matching the above lower bound.

Before we proceed to the lower bound for the Collision problem, we recall the following definition:

Definition 4.1. A function $f : A \to B$ is called r-to-1 is every element in the image of f has exactly r pre-images.

Now, let $\mathcal{A}$ be a quantum algorithm which decides the collision problem in t queries to $\mathcal{O}_w$, with probability $\geq 2/3$. Let $P(\widetilde{w})$ be the accepting polynomial of $\mathcal{A}$, as given by Corollary 4.3. We have $\deg(P) \leq 2t$, and for any $w \in [M]^N$, we have:

1. If w is injective, then $P(\widetilde{w}) \in [0, \frac{1}{3}]$.
2. If w is 2-to-1, then $P(\widetilde{w}) \in [\frac{2}{3}, 1]$.
3. For any w, then $P(\widetilde{w}) \in [0, 1]$. Note that even though the collision problem is a promise problem, running $\mathcal{A}$ on a w (not necessarily satisfying the promise) will still cause $\mathcal{A}$ to output a legitimate probability.

We now present an argument due to [Kut05] for proving the collision lower bound. Now, define:

Definition 4.2. Let $a, b, t \in \mathbb{N}$ be natural numbers such that $a \mid t, b \mid (N - t)$, where $t \leq N$. Such a tuple (t, a, b) is called valid. Then define:

$$W_{t,a,b} := \left\{ w \in [M]^N : \exists T \in \binom{[N]}{t}, w|_T \text{ is } a\text{-to-1}, w|_{[N]\setminus T} \text{ is } b\text{-to-1} \right\}$$

For example, consider the function $f : [10] \to [4]$ given by $f(1) = f(2) = 1, f(3) = f(4) = 2, f(5) = f(6) = f(8) = 3, f(7) = f(9) = f(10) = 4$. Then $f \in W_{4,2,3}$, as is witnessed by $T = \{1, 2, 3, 4\}$.
Also note that $W_{t,a,b} = W_{N-t,b,a}$. Finally, $\bigcup_{t=0}^{N} W_{t,1,1}$ is the set of all 1-to-1 functions. Similarly, $\bigcup_{t=0}^{N/2} W_{2t,2,2}$ is the set of all 2-to-1 functions.
Finally, define the symmetrization of P to be:

$$Q(t, a, b) := \mathop{\mathbb{E}}_{w \sim W_{t,a,b}} \left[P(\widetilde{w}) \right] = \frac{1}{|W_{t,a,b}|} \sum_{w \in W_{t,a,b}} P(\widetilde{w})$$

Note that $Q(t, 1, 1) \in [\frac{2}{3}, 1]$ for all $t \in [0, N] \cap \mathbb{Z}$, $Q(t, 2, 2) \in [0, \frac{1}{3}]$ for all $t \in [0, N] \cap 2\mathbb{Z}$, and $Q(t, a, b) \in [0, 1]$ for all valid tuples (t, a, b). We now claim that the degree of $Q(\cdot, \cdot, \cdot)$ is bounded by $\deg(P)$.

Lemma 4.9. There exists a polynomial $\mathcal{Q} : \mathbb{R}^3 \to \mathbb{R}$ such that for any valid tuple $(t, a, b) \in \mathbb{Z}^3 \subset \mathbb{R}^3$, $\mathcal{Q}(t, a, b) = Q(t, a, b)$. Furthermore, we can take $\deg(\mathcal{Q}) \leq \deg(P)$.

Proof. Refer [Kut05], Lemma 2.2. ∎

Remark. From now on we will conflate Q and $\mathcal{Q}$.

We shall need another result from approximation theory due to Paturi [Pat92].

Theorem 4.10. Let $q : \mathbb{R} \to \mathbb{R}$ be a polynomial of degree d. Let $a, b \in \mathbb{R}$ such that $a < b$, and let $\xi \in [a, b]$ be a real number. Suppose $\max_{x \in [a,b] \cap \mathbb{Z}} |q(x)| \leq c_1, c_2 := |q(\lceil \xi \rceil) - q(\xi)|$. Then

$$d = \Omega_{c_1, c_2} \left(\sqrt{(\xi - a + 1)(b - \xi + 1)} \right)$$

In particular, if $c_1, c_2 = \mathcal{O}(1)$, then $d = \Omega \left(\sqrt{(\xi - a + 1)(b - \xi + 1)} \right)$.

Set $M = 2\lfloor N/4 \rfloor$. Now we make cases:

1. $Q(M, 1, 2) \geq 1/2$: Write $g(x) := Q(M, 1, 2x)$, and let k be the least positive integer such that $|g(k)| \geq 2$. Then $\max_{x \in [0, k-1] \cap \mathbb{Z}} |g(x)| < 2$. Also, $g(1) - g(1/2) = Q(M, 1, 2) - Q(M, 1, 1) \geq \frac{1}{6}$. Thus invoking Theorem 4.10 (with $a = 0, b = k - 1, \xi = 1/2$), we get $\deg(g) = \Omega(\sqrt{k})$. Now consider the polynomial $h(x) := Q(N - 2kx, 1, 2k)$. For $x \in [0, N/2k] \cap \mathbb{Z}$, $(N - 2kx, 1, 2k)$ is a valid tuple and thus $h(x) \in [0, 1]$, i.e. $\max_{x \in [0, N/2k] \cap \mathbb{Z}} |h(x)| \leq 1$. However, $|h((N - M)/2k)| = |Q(M, 1, 2k)| = |g(k)| \geq 2$. Thus, once again invoking Theorem 4.10 (with $a = 0, b = N/2k, \xi = (N - M)/2k$) yields $\deg(h) = \Omega(N/k)$, where we use the fact that $M \approx N/2$. Now, if $k = \Omega(N^{2/3})$, then $\deg(g) = \Omega(\sqrt{k}) = \Omega(N^{1/3})$. However, $\deg(g) \leq \deg(Q) \leq \deg(P) \leq 2t$, and thus $t = \Omega(N^{1/3})$. Similarly, if $k = \mathcal{O}(N^{2/3})$, then $\deg(h) = \Omega(N/k) = \Omega(N^{1/3})$. However, $\deg(h) \leq \deg(Q) \leq \deg(P) \leq 2t$, and thus $t = \Omega(N^{1/3})$. Thus, in either case we have $t = \Omega(N^{1/3})$, as desired.

2. $Q(M, 1, 2) < 1/2$: Write $g(x) := Q(M, 2x, 2)$, and let k be the least positive integer such that $|g(k)| \geq 2$. Then $\max_{x \in [0, k-1] \cap \mathbb{Z}} |g(x)| < 2$. Also, $g(1) - g(1/2) = Q(M, 2, 2) - Q(M, 1, 2) \geq \frac{1}{6}$. Thus invoking Theorem 4.10 (with $a = 0, b = k - 1, \xi = 1/2$), we get $\deg(g) = \Omega(\sqrt{k})$. Now consider the polynomial $h(x) := Q(2kx, 2k, 2)$. For $x \in [0, N/2k] \cap \mathbb{Z}$, $(2kx, 2k, 2)$ is a valid tuple [24] and thus $h(x) \in [0, 1]$, i.e. $\max_{x \in [0, N/2k] \cap \mathbb{Z}} |h(x)| \leq 1$. However, $|h(M/2k)| = |Q(M, 2k, 2)| = |g(k)| \geq 2$. Thus, once again invoking Theorem 4.10 (with $a = 0, b = N/2k, \xi = M/2k$) yields $\deg(h) = \Omega(N/k)$. At this point, we're done as above.

Thus, to summarize, we have the following result:

Theorem 4.11 (Lower Bound for the Collision Problem)**.** Let $f : [N] \to [M]$ be a function, that is promised to be either injective or 2-to-1. Then any quantum algorithm, correctly deciding this problem with probability $\geq 2/3$, must necessarily make $\Omega(N^{1/3})$ queries to $\mathcal{O}_f$.

Remark. An exactly similar proof as above works to show that if f is promised to be injective or r-to-1, then the corresponding lower bound is $\Omega((N/r)^{1/3})$.

[24] WLOG we can assume N is even: Otherwise it will be impossible for f to be 2-to-1 and we can directly declare f to be injective in the beginning

References

[Aar16a] Scott Aaronson. The complexity of quantum states and transformations: From quantum money to black holes, 2016. URL: https://arxiv.org/abs/1607.05256, arXiv:1607.05256.

[Aar16b] Scott Aaronson. Introduction to quantum information science lecture notes. https://www.scottaaronson.com/qclec.pdf, 2016. [Online; accessed 27-July-2024].

[ABB+17] Andris Ambainis, Kaspars Balodis, Aleksandrs Belovs, Troy Lee, Miklos Santha, and Juris Smotrovs. Separations in query complexity based on pointer functions. *J. ACM*, 64(5), sep 2017. doi:10.1145/3106234.

[ABDK+21] Scott Aaronson, Shalev Ben-David, Robin Kothari, Shravas Rao, and Avishay Tal. Degree vs. approximate degree and quantum implications of huang's sensitivity theorem. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, pages 1330–1342, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3406325.3451047.

[AD11] Scott Aaronson and Andrew Drucker. Advice coins for classical and quantum computation. In *Proceedings of the 38th International Colloquim Conference on Automata, Languages and Programming - Volume Part I*, ICALP'11, pages 61–72, Berlin, Heidelberg, 2011. Springer-Verlag.

[AKS04] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. Primes is in p. *Annals of Mathematics*, 160(2):781–793, 2004. URL: http://www.jstor.org/stable/3597229.

[Amb07] Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM Journal on Computing*, 37(1):210–239, 2007. arXiv:https://doi.org/10.1137/S0097539705447311, doi:10.1137/S0097539705447311.

[Ber98] Daniel J. Bernstein. Detecting perfect powers in essentially linear time. *Math. Comput.*, 67(223):1253–1283, jul 1998. doi:10.1090/S0025-5718-98-00952-1.

[BS21] Nikhil Bansal and Makrand Sinha. k-forrelation optimally separates quantum and classical query complexity. STOC 2021, pages 1303–1316, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3406325.3451040.

[BV97] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. arXiv:https://doi.org/10.1137/S0097539796300921, doi:10.1137/S0097539796300921.

[DFG74] Antonio Degasperis, Luciano Fonda, and Giancarlo Ghirardi. Does the lifetime of an unstable system depend on the measuring apparatus? *Il Nuovo Cimento A (1965-1970)*, 21:471–484, 1974. URL: https://api.semanticscholar.org/CorpusID:120279111.

[DJ92] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439:553 – 558, 1992. URL: https://api.semanticscholar.org/CorpusID:121702767.

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96, pages 212–219, New York, NY, USA, 1996. Association for Computing Machinery. doi:10.1145/237814.237866.

[HC70] Martin E. Hellman and Thomas M. Cover. Learning with Finite Memory. *The Annals of Mathematical Statistics*, 41(3):765 – 782, 1970. doi:10.1214/aoms/1177696958.

[Kut05] Samuel Kutin. Quantum lower bound for the collision problem with small range. *Theory of Computing*, 1(2):29–36, 2005. URL: https://theoryofcomputing.org/articles/v001a002, doi:10.4086/toc.2005.v001a002.

[LMW24] Alex Lombardi, Fermi Ma, and John Wright. A one-query lower bound for unitary synthesis and breaking quantum cryptography. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, pages 979–990, New York, NY, USA, 2024. Association for Computing Machinery. doi:10.1145/3618260.3649650.

[OW15] Ryan O'Donnell and John Wright. Quantum computation scribe notes. https://www.cs.cmu.edu/~odonnell/quantum15/QuantumComputationScribeNotesByRyanODonnellAndJohnWright.pdf, 2015. [Online; accessed 27-July-2024].

[Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '92, pages 468–474, New York, NY, USA, 1992. Association for Computing Machinery. doi:10.1145/129712.129758.

[Sho94] Peter W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, 1994. URL: https://api.semanticscholar.org/CorpusID:15291489.

[Sim97] Daniel R. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997. arXiv:https://doi.org/10.1137/S0097539796298637, doi:10.1137/S0097539796298637.

[SSW23] Alexander A. Sherstov, Andrey A. Storozhenko, and Pei Wu. An optimal separation of randomized and quantum query complexity. *SIAM Journal on Computing*, 52(2):525–567, 2023. arXiv:https://doi.org/10.1137/22M1468943, doi:10.1137/22M1468943.

[Teu04] Christof Teuscher. *Alan Turing: Life and Legacy of a Great Thinker*. SpringerVerlag, 2004.

[Vaz04] Umesh Vazirani. Shor's factoring algorithm. https://people.eecs.berkeley.edu/~vazirani/f04quantum/notes/lec9.pdf, 2004. [Online; accessed 29-July-2024].

[Wik24] Wikipedia. Partial trace — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Partial%20trace&oldid=1225818151, 2024. [Online; accessed 27-July-2024].