SHOR'S ALGORITHM AND MATHEMATICAL FOUNDATIONS OF QUANTUM COMPUTING

BENJAMIN TANG

Abstract. Quantum computation provides efficient algorithms for certain computational problems that are believed to be difficult for classical computers. This paper develops the mathematical framework required to analyze quantum computation and applies it to Shor's factoring algorithm, analyzing period-finding and continued fractions. These results explain why Shor's algorithm factors integers in polynomial time on a quantum computer, while no polynomial time classical factoring algorithm is known.

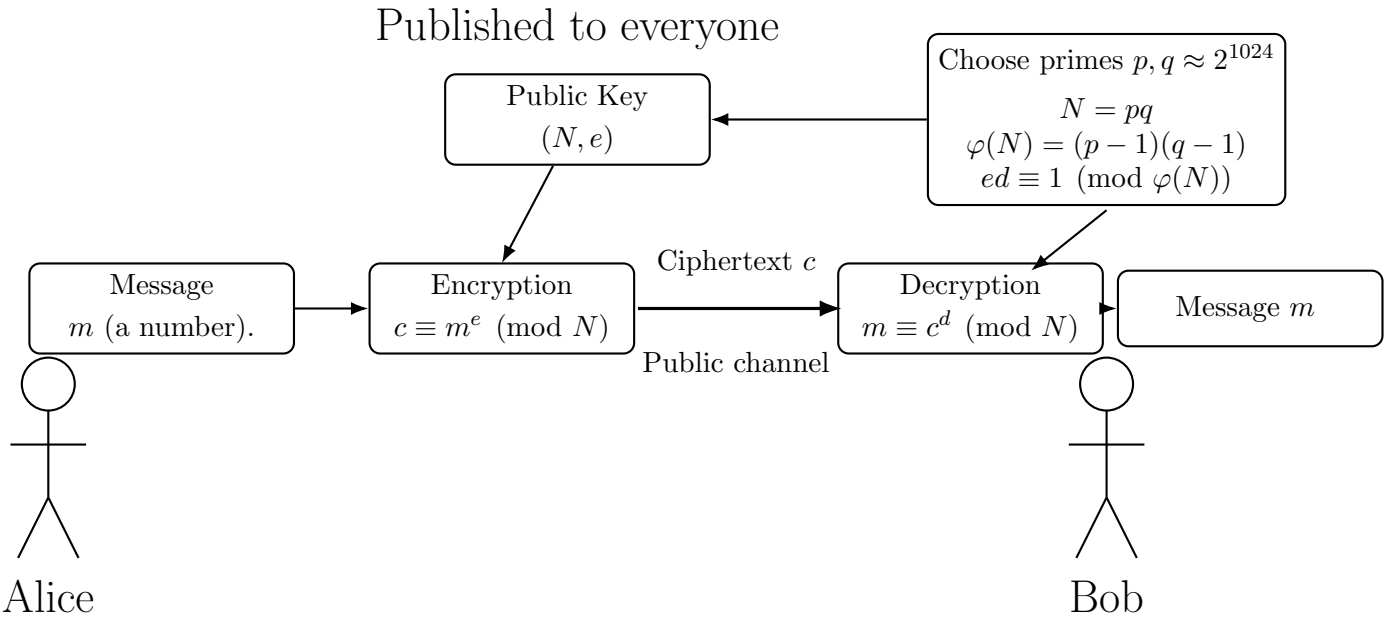
1. Introduction

Cryptography has become essential in our life, protecting our privacy, ensuring security, preventing theft, and more. One type of essential cryptography is a method for private communication over an insecure channel with someone you do not share a secret key with, also known as asymmetric encryption.

But how do you secretly send information to a stranger? There are many ways to solve this problem. We examine RSA encryption, which is used in a sizable proportion of asymmetric encryption today, including a part of HTTPS. Note that understanding RSA encryption is not essential for the content of this paper.

First, we examine how RSA encryption works. Suppose that Alice wants to send a message to Bob using RSA encryption. We assume Alice wishes to send a number. This easily generalizes to letters because letters can be converted to numerical form. RSA encryption is shown in the schematic below.

Date: June 8–July 10, 2026.



Note that p and q are approximately 1024 bits, so N is approximately 2048 bits. This corresponds to a 2048-bit RSA key.

The decryption method relies on Euler's Generalization of Fermat's Little Theorem, which states

Theorem 1.1 (Euler's Generalization of Fermat's Little Theorem)**.** *Let N be a natural number. If $a \in [1, N-1]$ with $\gcd(a, N) = 1$, then $a^{\varphi(N)} = 1 \pmod{N}$, where $\varphi(N)$ is the Euler-Totient function.*

Since m will almost always be relatively prime to N (Alice can explicitly check this), $(m^e)^d = m^{ed} = m^{k\varphi(N)+1} = m \pmod{N}$ as desired.

It is believed a hacker who wants to access the message Alice sends needs to be able to factor N into its two prime factors (this is still an open problem). But how do we factor N into a product of two primes? There are many ways to brute force this on a classical computer, but they all take exponential time in terms of the number of bits, b. The General Number Field Sieve (GNFS) is the fastest known way on a classical computer to factor N into a product of two primes, and does so in $O\left(\exp\left(\left(\sqrt[3]{\frac{64}{9}\ln 2} + o(1)\right) b^{1/3} (\ln b)^{2/3}\right)\right)$ time, which is subexponential in b. We will not prove this result, but it is shown in [BLP93]. Instead, we look at a simpler search algorithm. A naive approach would be to check every number $2, 3, \ldots, \lfloor\sqrt{N}\rfloor$. Each check requires us to see if $N = 0 \pmod{c}$ is true.

On a classical computer, the division algorithm to check the remainder runs in $O(\log N * \log c)$ time.

Hence, the brute force trial division algorithm takes $O(2^{\frac{b}{2}} b^2)$ time, or exponential time. Thus, for large b (e.g., $b = 2048$ used in RSA), this algorithm will take an incredibly long time to run.

This is where Shor's algorithm comes in. Shor's algorithm is a quantum algorithm invented by Peter Shor in 1994 that solves the factoring problem $N = pq$ using $O(b^3)$ quantum operations, where b is the number of bits of the prime number [Sho94]. This is substantially faster than classical factoring algorithms as Shor's algorithm runs in *polynomial* time instead of exponential time. This poses a great security risk to RSA encryption if a strong enough quantum computer is made. Even though, as of the current date, we are not even close, quantum computers are improving dramatically. Thus, the objective of this paper is to explain Shor's algorithm.

In the first section, we will go over standard definitions and notions in quantum computing. In the second section, we introduce impossibility theorems that show (partially) the limitations of quantum computing. Then, we introduce quantum gates used for algorithms in section four. Popular examples of quantum computing will be explained in section 5. Section 6 will introduce basic quantum algorithms. Finally, section 7 will explain Shor's algorithm.

2. Quantum Computing Fundamentals

To begin working with the qubit, we need to establish our space.

Definition 2.1. A **Hilbert space** $\mathcal{H}$ is a vector space for quantum states with an inner product. For Hilbert spaces with finite dimensions, vectors are elements in $\mathbb{C}^n$.

For this paper, we will only study Hilbert spaces with finite dimensions.

Definition 2.2. A pure quantum state $|\psi_a\rangle$ called a **ket** lives in the Hilbert space $\mathcal{H}_a$ as a linear superposition of the orthonormal basis of the Hilbert space with coefficients $\in \mathbb{C}$.

Example. $|\psi_a\rangle \in \mathcal{H}_a$, $|\psi_a\rangle = \frac{i}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle = \begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$.

Notably, the statement "pure quantum" state rather than "quantum state" are not interchangeable. There are fundamental differences between them that we will analyze later.

To establish the inner product of this Hilbert space, we analyze the bra.

Definition 2.3. A **bra** $\langle\psi|$ is the conjugate transpose, or dagger, of a ket. Alternatively, it is written as a row vector.

Example. $\langle\psi| = \frac{-i}{\sqrt{2}}\langle 0| + \frac{1}{\sqrt{2}}\langle 1| = \begin{pmatrix} -\frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$.

Now we can define our inner product.

Definition 2.4. The **inner product** of pure quantum states in the Hilbert space is defined as $\langle\psi|\phi\rangle = \psi^\dagger\phi$.

Example. $|\psi\rangle = \frac{(i|0\rangle+|1\rangle)}{\sqrt{2}}$.

$\langle\psi|\psi\rangle = \frac{(i|0\rangle+|1\rangle)}{\sqrt{2}}(\frac{(-i\langle 0|+\langle 1|)}{\sqrt{2}}) = 1.$

$\psi^\dagger\psi = \begin{pmatrix} \frac{-i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = 1.$

Note that this is a real number when $|\psi\rangle = |\phi\rangle$, but is not necessarily a real number when $|\psi\rangle \neq |\phi\rangle$.

Now, the inner product is fundamental to the collapse of the wave function. In fact, the coefficient squared is the probability the wavefunction will collapse to the basis state. This can be formalized with the following postulate.

Postulate 2.5 (Born Rule)**.** *Let $\psi \in \mathcal{H}$ and e_i be a basis of $\mathcal{H}$. Then* $\mathbb{P}(\psi = e_i) = |\langle e_i|\psi\rangle|^2$.

In other words, if a quantum state is measured in its computational basis, the wavefunction collapses and the probability of being a certain state is the absolute value of the coefficient of that state squared.

Notice that we stated this was a postulate, not a theorem. This is because this statement cannot actually be derived from the fundamental postulates of quantum mechanics for the case $n = 2$ (the Hilbert space has two basis vectors). However, [Gle57] proved that the born rule is true for any Hilbert space with more than 2 basis vectors. We actually need to accept the $n = 2$ case as an axiom!

The following corollary, derived from the probability summing up to 1, is very useful.

Corollary 2.6. *A pure quantum state $|\psi\rangle$ must satisfy $\langle\psi|\psi\rangle = 1$.*

Note: global phases of a quantum state aren't essential because they don't affect measurements. However, local phases matter during entanglement. This can be made fully rigorous when we talk about density matrices later in this section.

Now, we want to formalize our idea of the measurement.

Definition 2.7. A matrix H is called **Hermitian** if and only if it satisfies $H^\dagger = H$.

Example. $\begin{pmatrix} \frac{1}{\sqrt{2}} & -2i \\ 2i & 3 \end{pmatrix}$ is Hermitian.

Although this definition may seem uncalled for, it is actually essential to our understanding of the measurement. But what even is a measurement?

Postulate 2.8. *An observation is mathematically equivalent to multiplying the bra on the left and the ket on the right, with anything in between.*

Theorem 2.9 (Uniqueness)**.** *If $\langle\psi| X |\psi\rangle = 0$ for all $|\psi\rangle$, then $X = 0$.*

Proof. We can write $X = \sum_{i=0}^{n}\sum_{j=0}^{n} \alpha_{i,j} |\epsilon_i\rangle \langle\epsilon_j|$.

First, let $|\psi\rangle = |\epsilon_i\rangle$ for some i.

Then we have $\alpha_{i,i} = 0$ for all i.

Next, let $|\psi\rangle = \gamma|\epsilon_i\rangle + |\epsilon_j\rangle$ for some i, j (normalization isn't required because 0 remains invariant among multiplication).

Then, $\langle\psi| X |\psi\rangle = \alpha_{i,j}\bar{\gamma} + \alpha_{j,i}\gamma = 0$.

Then, letting $\gamma = 1, i$ we have

$$\begin{cases} \alpha_{i,j} + \alpha_{j,i} = 0 \\ \alpha_{i,j} - \alpha_{j,i} = 0 \end{cases}.$$

Hence, $\alpha_{i,j} = 0$ for all i, j. Thus, $X = 0$ as desired. ■

Since measurements are real, we can now get our theorem on what measurements we can make.

Theorem 2.10. *For a matrix H to represent a physically measurable quantity of a quantum state $|\psi\rangle$, it must be Hermitian.*

Proof. Observable quantities are real. Hence $\langle\psi|H|\psi\rangle = \langle\psi|H|\psi\rangle^\dagger = \langle\psi|H^\dagger|\psi\rangle$.

Since ψ is arbitrary, $H = H^\dagger$, which is the definition of a Hermitian matrix. ■

Example. Suppose we have a quantum state $|\psi\rangle = \frac{(|0\rangle+|1\rangle)}{\sqrt{2}}$.

The Hermitian matrix $H = |0\rangle\langle 0| - |1\rangle\langle 1|$ measures the state of ψ, returning 1 if $|\psi\rangle$ collapses $|0\rangle$, and -1 if $|\psi\rangle$ collapses to $|1\rangle$.

If we let $|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}$ and $|1\rangle = \begin{pmatrix}0\\1\end{pmatrix}$, then we have $H = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}$.

Now, pure quantum states aren't really useful if we can't change them. This motivates the definition below.

Definition 2.11. A matrix U is called **Unitary** if and only if $U^\dagger = U^{-1}$.

Example. $\frac{1}{\sqrt{2}}\begin{pmatrix}1 & i\\i & 1\end{pmatrix}$ is Unitary.

With our definition of Unitary, we can now operate on quantum states using the theorem below and a postulate!

Theorem 2.12. *If a matrix U can operate on any quantum state $|\psi\rangle$ in Hilbert space $\mathcal{H}$, it must be unitary.*

Proof. $1 = \langle U\psi|U\psi\rangle = \langle\psi|U^\dagger U|\psi\rangle \iff \langle\psi|\,U^\dagger U - I\,|\psi\rangle = 0$. Since ψ is arbitrary, $U^\dagger U = I$, or $U^\dagger = U^{-1}$, which is unitary by definition 2.11.

■

Postulate 2.13. *A matrix U can operate on a quantum state $|\psi\rangle$ if it is unitary.*

Thus, we can operate on quantum states!

Theorem 2.14. *Unitary matrices preserve the inner product of any two quantum states.*

Proof. Let U be a unitary matrix and let $|\psi\rangle$, $|\phi\rangle$ be quantum states. Then $\langle\psi|\phi\rangle = \langle\psi|(U^\dagger U)|\phi\rangle = \langle(U\psi)|(U\phi)\rangle$ as desired. ■

This might remind you of something. A matrix that preserves the inner product of any two quantum states is analogous to a matrix that preserves the dot product of any two vectors, which is simply a group of rigid rotations and reflections!

This motivates a definition called the **Bloch Sphere**, in which a unitary is a rotation on this sphere, along with adding a global phase. This will not be covered in this paper as it is not essential to our algorithms.

Now, there are some even more nice properties of unitary matrices. As we know, the O_n group has orthogonal columns (by definition). The same can be said for unitary matrices! To prove this, we use some notation.

Definition 2.15. We define $\delta_{i,j} = \begin{cases}1 & \text{if } i = j\\0 & \text{if } i \neq j\end{cases}$. This is known as the **Kronecker delta**. We add a comma for clarity (this is different from conventional notation).

Theorem 2.16. *Let the ith column of a unitary U be C_i. Then $\langle C_i|C_j\rangle = \delta_{i,j}$*

Proof. Let $\epsilon_{i,j}$ denote the ith row and jth column of U, where U is $N \times N$.

Let $a_{i,j}$ denote the coefficient of the ith row and jth column of U.

$$I = UU^\dagger = (\sum_{i=1}^{N}\sum_{j=1}^{N} a_{i,j}\epsilon_{i,j})(\sum_{i=1}^{N}\sum_{j=1}^{N} \bar{a}_{j,i}\epsilon_{i,j}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a_{k,i}\bar{a_{k,j}}\epsilon_{i,j} = \sum_{i=1}^{N}\sum_{j=1}^{N} \delta_{i,j}\epsilon_{i,j}.$$

Now, $C_i = \sum_{k=1}^{N} a_{k,i}\epsilon_{k,i}$.

$\langle C_i|C_j\rangle = \sum_{k=1}^{N} a_{k,j}\bar{a_{k,i}} = \delta_{j,i} = \delta_{i,j}$ as desired. ■

Truthfully, the proof listed above does not help with understanding and is simply there for rigor. An easier intuitive "proof" is noting that the condition $UU^\dagger = I$ is the same condition as the orthogonal columns when writing the multiplication out.

Now, we've only been talking about pure quantum states. However, this can give the reader a false impression that there only exists pure quantum states, which is anything but the truth. Thus, to formalize the definition of mixed (not pure) quantum states, we introduce the density matrix.

Definition 2.17. A **density matrix** ρ_ψ of pure state $|\psi\rangle$ is $|\psi\rangle\langle\psi|$.

Example. Let $|\psi\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}} = \begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}$. $\rho_\psi = \begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$.

This may be a bit uncalled for, but is of quintessential importance to measurement and defining decoherence.

Now, to show why density matrices are useful for measurement, we recall a fact from linear algebra.

Theorem 2.18. *Let A be a $n \times m$ matrix, and let B be a $m \times n$ matrix. Then* $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

Theorem 2.19. *For density matrix ρ of (not necessarily pure) quantum state $|\psi\rangle$, we have* $\mathrm{Tr}(\rho) = 1$.

Proof. $\mathrm{Tr}(\rho) = \mathrm{Tr}(|\psi\rangle\langle\psi|) = \mathrm{Tr}(\langle\psi|\psi\rangle) = \mathrm{Tr}(1) = 1$. ■

Theorem 2.20. *For density matrix ρ of pure state $|\psi\rangle$, we have* $\mathrm{Tr}(\rho^2) = 1$.

Proof. $\mathrm{Tr}\left((|\psi\rangle\langle\psi|)^2\right) = \mathrm{Tr}(|\psi\rangle\langle\psi|\psi\rangle\langle\psi|) = \mathrm{Tr}(|\psi\rangle\langle\psi|) = 1$. ■

Notice I stated *pure state* $|\psi\rangle$ instead of quantum state $|\psi\rangle$ in the proof above. That is because it's not actually true for mixed states!

Theorem 2.21. *For hermitian matrix H, we have* $\langle\psi|H|\psi\rangle = \mathrm{Tr}(\rho_\psi H)$.

Proof.

$$\begin{aligned}\mathrm{Tr}(\rho_\psi H) &= \mathrm{Tr}(|\psi\rangle\langle\psi| H) \\ &= \mathrm{Tr}(\langle\psi| H |\psi\rangle) \\ &= \langle\psi|H|\psi\rangle .\end{aligned}$$

Since $\langle\psi|H|\psi\rangle \in \mathbb{R}$. ■

Theorem 2.22. *Let ρ be the density matrix of quantum state $|\psi\rangle$, and let U be unitary. Then, the density matrix of $|U\psi\rangle$ is $U\rho U^{-1}$.*

Proof. $\rho = |\psi\rangle\langle\psi|$.
$|U\psi\rangle\langle U\psi| = U|\psi\rangle\langle\psi|U^{-1} = U\rho U^{-1}$ as desired. ■

Thus, ANY measurement and unitary can be written as an operation on a density matrix. This shows why the density matrix is so fundamental!

Next, we talk about decoherence.

Definition 2.23. A **mixed quantum state** is a statistical ensemble of pure states (ex: it was measured). Mixed quantum states CANNOT be written as one entire wave function because they would not be normalized.

Remark 2.24. One cannot write a mixed quantum state as $\sum p_i |\psi_i\rangle$ because this notation allows for amplitude addition. For example, if $|\psi_1\rangle = |0\rangle$ and $|\psi_2\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$, we can't add the $|0\rangle$ and $|1\rangle$ components together separately. Thus, this motivates us to use the density matrix to express mixed states.

Theorem 2.25. *The density matrix of a mixed quantum state is $\sum p_i |\psi_i\rangle\langle\psi_i|$, where $\sum p_i = 1$. p_i means the probability of the quantum state as $|\psi_i\rangle$. One can explicitly check the above theorems on $\mathrm{Tr}(\rho) = 1$, measurement operations, and unitary operators still apply.*

What is really nice is that we can now distinguish a mixed quantum state with a pure quantum state.

Theorem 2.26. *If ρ is the density matrix of a mixed state $|\psi\rangle$, then*
$\mathrm{Tr}(\rho^2) < 1$.

Proof. By definition of a mixed state density matrix, we have

$$\begin{aligned}\mathrm{Tr}(\rho^2) &= \mathrm{Tr}((\sum p_i |\psi_i\rangle\langle\psi_i|)^2) \\ &= \mathrm{Tr}(\sum p_i^2 |\psi_i\rangle\langle\psi_i| + \sum_{i\neq j} p_i p_j |\psi_i\rangle\langle\psi_i|\psi_j\rangle\langle\psi_i|) \\ &= \sum p_i^2 + \sum_{i\neq j}(p_i p_j \langle\psi_i|\psi_j\rangle^2) \\ &= (\sum p_i)^2 - \sum_{i,j} p_i p_j (1 - \langle\psi_i|\psi_j\rangle^2) \\ &= (1 - \sum_{i,j}(p_i p_j (1 - \langle\psi_i|\psi_j\rangle^2))) < 1\end{aligned}$$

as desired. ■

Notice that our Born rule postulate only applied to pure quantum states. We now apply it to mixed quantum states.

Postulate 2.27 (Extension Of Born Rule To Mixed Quantum States)**.** *Let O be a Hermitian matrix, diagonalized to $\sum V\Lambda V^\dagger = \sum_i \lambda_i |\epsilon_i\rangle \langle\epsilon_i|$.*
Let ρ be a density matrix. Then
$\mathbb{P}(O = \lambda) = \mathrm{Tr}(\sum_{\lambda_i=\lambda} \rho |\epsilon_i\rangle \langle\epsilon_i|)$.

This fundamentally states the same idea as the born rule.
We now wish to introduce entanglement. To do that, we establish the tensor product.

Definition 2.28. The tensor product $\mathcal{H}_a \otimes \mathcal{H}_b$ is the Hilbert space generated by vectors of the form $|\psi\rangle_a \otimes |\phi\rangle_b$.

Example. $\frac{|0\rangle_a+|1\rangle_a}{\sqrt{2}} \in \mathcal{H}_a, |0\rangle_b \in \mathcal{H}_b, (\frac{|0\rangle_a+|1\rangle_a}{\sqrt{2}}) \otimes |0\rangle_b = \frac{|0\rangle_a|0\rangle_b+|1\rangle_a|0\rangle_b}{\sqrt{2}} \in \mathcal{H}_a \otimes \mathcal{H}_b$.

What happens to density matrices in a tensor product is also of interest to us.

Theorem 2.29. *Suppose the density matrix of $|\alpha\rangle$ is ρ, and the density matrix of $|\beta\rangle$ is q. Then the density matrix of $|\alpha\rangle \otimes |\beta\rangle$ is $\rho \otimes q$*

Proof. The density matrix of $|\alpha\rangle \otimes |\beta\rangle$ is $(|\alpha\rangle \otimes |\beta\rangle)(\langle\alpha| \otimes \langle\beta|) = \rho \otimes q$.
Where we used the identity $(A \otimes C)(B \otimes D) = AB \otimes CD$. ■

Remark 2.30. $|\alpha\rangle \otimes |\beta\rangle = |\alpha, \beta\rangle = |\alpha\beta\rangle$.

Definition 2.31 (Qubit)**.** A **qubit** is a quantum system whose Hilbert space is two-dimensional (it can be either $|1\rangle$ or $|0\rangle$, like how you would expect a bit to be).

Now we can introduce entanglement!

Definition 2.32 (Bell States)**.** We define the bell states as

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

Notice that for these states, measuring one qubit in its computational basis determines the measurement of the other qubit in its computational basis. These states are **maximally entangled**.

Theorem 2.33 (Partial Density Matrix)**.** *Suppose $|ab\rangle \in \mathcal{H}_a \otimes \mathcal{H}_b$.*
Let the density matrix ρ_{ab} of $|ab\rangle$ be $\sum c_i |\epsilon^i_{a1}\epsilon^i_{b1}\rangle \langle\epsilon^i_{a2}\epsilon^i_{b2}|$, where ϵ are the basis states.
Then $\rho_a = \sum c_i |\epsilon^i_{a1}\rangle \langle\epsilon^i_{a2}| \delta_{b1,b2}$.

Proof. If we perform the measurement $H_a \otimes I_b$ on ρ_{ab} for Hermitian matrix H_a, the expectation value should be the same as performing the measurement H_a on ρ_a because measuring identity returns 1. Therefore

$\text{Tr}((H_a \otimes I_b)\rho_{ab}) = \text{Tr}(H_a\rho_a)$ must be true. We claim this can only be true for our desired value of ρ_a.

$$\begin{aligned}\text{Tr}((H_a \otimes I_b)\rho_{ab}) = \text{Tr}((H_a \otimes I_b)\sum c_i \left|\epsilon^i_{a1}\epsilon^i_{b1}\right\rangle \left\langle\epsilon^i_{a2}\epsilon^i_{b2}\right|) = \\ \sum (c_i \,\text{Tr}(\langle\epsilon^i_{a1}|H_a|\epsilon^i_{a2}\rangle)\delta_{b1,b2}) = \\ \text{Tr}(H_a \sum (c_i \left|\epsilon^i_{a1}\right\rangle \left\langle\epsilon^i_{a2}\right|)\delta_{b1,b2}) = \\ \text{Tr}(H_a\rho_a).\end{aligned}$$

For this to be true for all hermitian matrices H_a, we must have $\rho_a = \sum c_i \left|\epsilon^i_a\right\rangle \left\langle\epsilon^i_a\right| \delta_{b1,b2}$, as desired.

■

Definition 2.34 (Entanglement)**.** Two pure quantum states $|a\rangle$ and $|b\rangle$ are **entangled** if and only if the trace of the reduced density matrix ρ_a squared ($\text{Tr}(\rho_a^2)$) is less than 1.

Intuitively, this means that measuring the state of one quantum state gives you some information about the other (this includes phases!).

3. Impossibility Theorems

Theorem 3.1 (No Cloning)**.** *There exists no unitary operator U acting on $\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}$ such that $U(|\psi\rangle \otimes |c\rangle \otimes |a\rangle) = |\psi\rangle \otimes |\psi\rangle \otimes |a'\rangle$ for all $\psi \in \mathcal{H}$.*

Proof. Assume such U exists.

Choose two $|\psi\rangle$, $|\phi\rangle$ such that $\langle\psi|\phi\rangle \neq 0, 1$

$$\begin{aligned}\langle\psi|\phi\rangle &= (\langle\psi| \otimes \langle c| \otimes \langle a|)(|\phi\rangle \otimes |c\rangle \otimes |a\rangle) \\ &= (\langle\psi| \otimes \langle c| \otimes \langle a|)U^\dagger U(|\phi\rangle \otimes |c\rangle \otimes |a\rangle) \\ &= (\langle\psi| \otimes \langle\psi| \otimes \langle a'|)(|\phi\rangle \otimes |\phi\rangle \otimes |a''\rangle) \\ &= \langle\psi|\phi\rangle^2 \langle a'|a''\rangle \implies \langle\psi|\phi\rangle = 0, 1.\end{aligned}$$

Since the inner product is bounded between 0 and 1, this is a contradiction. ■

No cloning has profound implications on quantum computing. For example, it limits the amount of information carried by a qubit because we cannot create multiple copies of the qubit and see the probability they collapse to certain states. This theorem also has a dual.

Theorem 3.2 (No Deleting)**.** *There exists no unitary operator U acting on $\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}$ such that $U(|\psi\rangle \otimes |\psi\rangle \otimes |a\rangle) = |\psi\rangle \otimes |0\rangle \otimes |a'\rangle$ for all $\psi \in \mathcal{H}$.*

Proof. Unitary operations are reversible. Hence, if this was possible then cloning is possible, which is a contradiction. ■

Next, we show that we cannot have faster than light communication with qubits. That is, a wave function collapse happens instantaneously across distances, but no information travels faster than the speed of light.

Theorem 3.3 (No Communicating)**.** *Suppose $|ab\rangle \in \mathcal{H}_a \otimes \mathcal{H}_b$. Then no local unitary or measurement on $|b\rangle$ changes the reduced density matrix ρ_a.*

Proof. Let the density matrix ρ_{ab} of $|ab\rangle$ be $\sum c_i \left|\epsilon^i_{a1}\epsilon^i_{b1}\right\rangle \left\langle\epsilon^i_{a2}\epsilon^i_{b2}\right|$, where ϵ are the basis states.

We know $\rho_a = \sum c_i \left|\epsilon^i_{a1}\right\rangle \left\langle\epsilon^i_{a2}\right| \delta_{b1,b2}$.

If we perform a unitary on $|b\rangle$, then the new joint density matrix becomes

$\rho'_{ab} = \sum c_i \left|\epsilon^i_{a1}(U\epsilon^i_{b1})\right\rangle \left\langle\epsilon^i_{a2}(U\epsilon^i_{b2})\right|$.

However, $\delta_{b_1,b_2} = \delta_{\epsilon_{b_1},\epsilon_{b_2}} = \delta_{U(\epsilon_{b_1}),U(\epsilon_{b_2})}$, so the reduced density matrix ρ_a doesn't change.

If we perform a measurement on $|b\rangle$, then all the off diagonal terms are lost (assuming measurement in the computational basis. However, a measurement in any basis can be achieved by a unitary and then a measurement in the computational basis), and we have

$\rho'_{ab} = \sum c_i \left|\epsilon^i_{a1}(\epsilon^i_{b1})\right\rangle \left\langle\epsilon^i_{a2}(\epsilon^i_{b2})\right| \delta_{b_1,b_2}$.

By inspection ρ_a does not change. Thus, we have proved no communicating.

■

4. Quantum Gates

To start quantum algorithms, we need to introduce quantum gates.

Definition 4.1 (Swap Gate)**.** A swap gate operating on $\mathcal{H}_a \otimes \mathcal{H}_b$ is the unitary $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

Intuitively, it swaps the qubits $|a\rangle$ and $|b\rangle$.

Definition 4.2 (Hadamard Gate)**.** A **Hadamard gate** is the primary gate used to establish superposition. It is represented by unitary $H = \frac{\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}{\sqrt{2}}$.

It does the following on basis states.

$|0\rangle \xrightarrow{H} \frac{|0\rangle+|1\rangle}{\sqrt{2}}$,

$|1\rangle \xrightarrow{H} \frac{|0\rangle-|1\rangle}{\sqrt{2}}$.

We can generalize the Hadamard gate to multiple qubits.

Definition 4.3 (Hadamard Gate With N Qubits)**.** A Hadamard gate with N Qubits, or $H^{\otimes N}$, does the following on basis states (note the value $|k\rangle$ ranges from 0 to $2^N - 1$, and converted to binary represents the state of the N qubits).

$|k\rangle \xrightarrow{H^{\otimes N}} \frac{1}{2^{\frac{N}{2}}} \sum_{i=0}^{2^N-1} (-1)^{k\cdot i} |i\rangle$

where $k \cdot i$ denotes the dot product when k and i are written in binary (e.g., $7 \cdot 2 = (111)_2 \cdot (010)_2 = 0 + 1 + 0 = 1$.

We leave it as an exercise to the reader to prove that this gate is indeed $H^{\otimes n}$.

Now that we have a gate to establish superposition, the next step is to be able to establish entanglement.

Definition 4.4 (CNOT Gate)**.** Suppose $|a\rangle$, $|b\rangle$ are qubits. The CNOT gate with the control $|a\rangle$ flips the $|b\rangle$ bit if and only if the $|a\rangle$ bit is $|1\rangle$. It is written as the matrix $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$

operating on $|a\rangle \otimes |b\rangle$.

Another important quantum gate is the **Toffoli gate**. We will not cover this gate in this paper. However, it is foundational to establish that any algorithm on a classical computer can be converted onto a quantum computer, because Toffoli gates are complete regarding an ancillary (meaning we can create any classical logic gate with them).

Definition 4.5 (Pauli Gates)**.** The Pauli Gates X, Y, Z as $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

One important advantage in quantum computing is that the wave-like behavior of quantum states allow for constructive and destructive interference. To take advantage of this, we use gates relating to phase shifts.

Definition 4.6 (Phase Shift Gate)**.** We define phase shift gate $R_\phi = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}$.

Intuitively, this shifts the phase corresponding to the $|1\rangle$ state by ϕ.

Definition 4.7 (Controlled Phase Gate)**.** We defined the controlled phase gate to be $C(\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\phi} \end{bmatrix}$.

That is, we will apply a phase shift gate of phase ϕ if and only if the control is $|1\rangle$.

All the gates defined above (besides $H^{\otimes n}$) is the quantum gate complexity of **one** quantum gate. Therefore, to find the quantum gate complexity of an arbitrary unitary, you need to decompose the unitary into basic gates and count how many gates there will be.

Alas, we define the quantum Fourier transform, which is very similar to the discrete Fourier transform.

Definition 4.8 (Quantum Fourier Transform And Inverse Quantum Fourier Transform)**.** Let $2^n = N$. We define the quantum Fourier transform on n qubits, or QFT, to do the following on basis states.

$|j\rangle \to \frac{1}{2^{\frac{n}{2}}} \sum_{k=0}^{2^n-1} e^{\frac{2\pi i j k}{N}} |k\rangle$. One can check that this operation is unitary.

The inverse quantum Fourier transform, or QFT^{-1} does the following on basis states.

$|j\rangle \to \frac{1}{2^{\frac{n}{2}}} \sum_{k=0}^{2^n-1} e^{-\frac{2\pi i j k}{N}} |k\rangle$. One can check that applying QFT and then QFT^{-1} returns identity.

Theorem 4.9. *The QFT gate (and also the QFT^{-1} if we take the dagger) operating on N qubits take at most $O(N^2)$ basic quantum gates.*

Proof. We claim that the following series of quantum gates is equivalent to the QFT.

Let $f(x)$ be a function returning a series of quantum gates, with x denoting the xth qubit. Define $f(x)$ as

(1) A Hadamard on the xth qubit.
(2) For every qubit k that satisfies $x < k \leq N$, a controlled phase gate with phase $\phi = \frac{2\pi i}{2^{k-x+1}}$ operating on qubit x, with qubit k as the control.

Now, we claim the QFT gate is equivalent to $f(1)f(2)f(3)\dots f(n)$ followed by swap gates with the k and the $n-k+1$ qubit for all $k \in [0, \lfloor \frac{n}{2} \rfloor]$. This takes $O(N^2)$ quantum gates by elementary combinatorics .

We now prove that this is equivalent to the QFT gate.

Let the nth qubit be in state $|x_n\rangle$ initially. We assume it is not in superposition (it is either $|0\rangle$ or $|1\rangle$).

Applying our operation (which we want to prove is the QFT) turns $\bigotimes_{i=1}^{N} |x_i\rangle \overset{?=QFT}{\to} \frac{1}{2^{\frac{N}{2}}} \bigotimes_{i=1}^{N} (|0\rangle + e^{2\pi i[0.x_{n-i+1}x_{n-i+2}x_{n-i+3}\dots x_n]} |1\rangle)$ where $0.abcd\dots$ denotes a decimal in binary.

Then, we have

$$\begin{aligned}
\frac{1}{2^{\frac{N}{2}}} \bigotimes_{i=1}^{N} (|0\rangle + e^{2\pi i[0.x_{n-i+1}x_{n-i+2}x_{n-i+3}\dots x_n]} |1\rangle) &= \frac{1}{2^{\frac{N}{2}}} \bigotimes_{i=1}^{N} (|0\rangle + e^{2\pi i[x_1x_2x_3\dots x_{n-i}.x_{n-i+1}x_{n-i+2}x_{n-i+3}\dots x_n]} |1\rangle) \\
&= \frac{1}{2^{\frac{N}{2}}} \bigotimes_{i=1}^{N} (|0\rangle + e^{\frac{2\pi i x 2^{N-i}}{2^N}} |1\rangle) \\
&= \frac{1}{2^{\frac{N}{2}}} \sum_{j=0}^{2^N-1} e^{\frac{2\pi j x}{2^N}} |j\rangle
\end{aligned}$$

as desired. ■

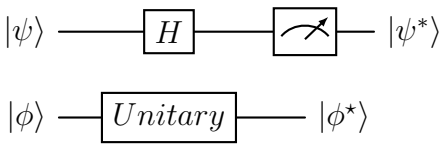
5. Circuits And Basic Quantum Algorithms

5.1. **Drawing A Circuit.** To draw a quantum circuit, several rules are applied.

Rule 5.1.1. *The initial quantum state starts on the left and the final quantum state is on the right. Thus, the algorithm goes from left to right.*

Rule 5.1.2. *Rectangles represent a local unitary or measurement on the quantum state. H, X, Y, Z represent the Hadamard and Pauli gates, respectively. A meter icon represents the measurement in the quantum state's computational basis.*

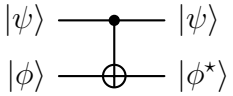
Example.



Where the $|\psi^\star\rangle$ state decohered.

Rule 5.1.3. *A CNOT gate is generally written as a shaded dot on the quantum state indicating the control, and an XOR symbol on the quantum state getting flipped. A line connects the two.*

Example.



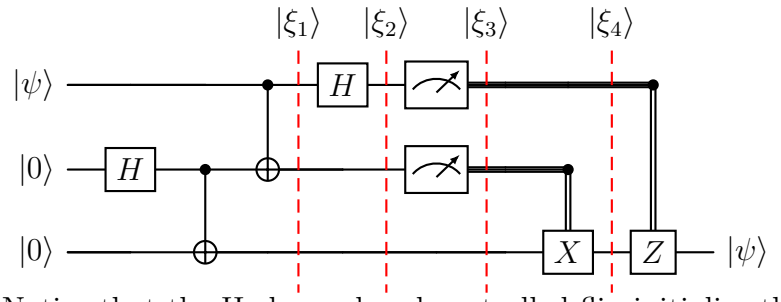
Rule 5.1.4 (Extension Of Rule 5.1.3)**.** *A shaded dot always indicates a control of any operation. Multiply qubits can do a control on another qubit. The operation only runs if all of the controls are* 1*. Specific to our paper,* U^C *denotes the unitary operation performed* C *times, where* C *is the value of the control.*

Rule 5.1.5. *A classical wire has two lines, unlike a quantum wire which has one line. They both mean a control, but a classical wire means that the quantum state was measured, and the information is sent classically.*

5.2. **Quantum Teleportation.** Quantum teleportation is a quantum algorithm that solves the following problem. Alice and Bob share a pair of maximally entangled qubits, and are in different locations. Suppose Alice has a qubit, and wants to give it to Bob without physically moving it (because this can risk decoherence). How does she do so?

The algorithm to accomplish such is represented by the circuit diagram below.

Let $|\xi\rangle$ denote the system state.



Notice that the Hadamard and controlled flip initialize the maximally entangled state.

Let $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$.

$|\xi_1\rangle = \frac{\alpha|000\rangle+\alpha|011\rangle+\beta|010\rangle+\beta|001\rangle}{\sqrt{2}}$.

$|\xi_2\rangle = \frac{\alpha(|100\rangle+|000\rangle+|011\rangle+|111\rangle)+\beta(|010\rangle+|001\rangle-|101\rangle-|110\rangle)}{2}$

Now, four different measurements can result for $|\xi_3\rangle$. Let the measurement of the first two qubits result in $|ab\rangle$. Then, let $|\xi_3\rangle_k$ denote the system state if the first two qubits resulted in the number k written in binary.

$|\xi_3\rangle_0 = |00\rangle \otimes (\alpha\,|0\rangle + \beta\,|1\rangle)$.

$|\xi_3\rangle_1 = |01\rangle \otimes (\alpha\,|1\rangle + \beta\,|0\rangle)$.

$|\xi_3\rangle_2 = |10\rangle \otimes (\alpha\,|0\rangle - \beta\,|1\rangle)$.

$|\xi_3\rangle_3 = |11\rangle \otimes (\alpha\,|1\rangle - \beta\,|0\rangle)$.

Then, for $|\xi_4\rangle$, we have

$|\xi_4\rangle_0 = |00\rangle \otimes (\alpha\,|0\rangle + \beta\,|1\rangle)$.

$|\xi_4\rangle_1 = |01\rangle \otimes (\alpha\,|0\rangle + \beta\,|1\rangle)$.

$|\xi_4\rangle_2 = |10\rangle \otimes (\alpha\,|0\rangle + \beta\,|1\rangle)$.

$|\xi_4\rangle_3 = |11\rangle \otimes (\alpha\,|0\rangle + \beta\,|1\rangle)$.

In all scenarios, the third qubit is now $|\psi\rangle$, as desired (teleportation was achieved).

5.3. **Deutsch's Algorithm.** In Deutsch's algorithm, a function $f : \{0,1\} \to \{0,1\}$ is used as quantum oracle O_w such that $|k\rangle \overset{O_w}{\to} (-1)^{f(k)}\,|k\rangle$.

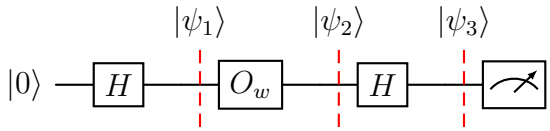
Remark 5.1. A more natural characterization of the quantum oracle O_w is $|k,l\rangle \overset{O_w}{\to} |k, l \otimes f(k)\rangle$. However, if we define a new quantum oracle $O'_w = (I \otimes H)O_w(I \otimes H)$, then the quantum oracle becomes $|k,l\rangle \overset{O'_w}{\to} (-1)^{lf(k)}\,|k,l\rangle$. If we let $|l\rangle = |1\rangle$, then we have the oracle stated above.

Our goal is to find $f(0) \oplus f(1)$. A classical computer finds this deterministically with two calls to the oracle.

However, notice that the classical computer got extra information. We don't need to know $f(0)$ to find $f(0) \oplus f(1)$, but the classical computer will know it. This suggests a quantum computer may be able to provide an advantage because it will only output the necessary information.

Deutsch's algorithm solves the problem deterministically with ONE oracle inquiry. The algorithm is shown below.

Exercise. Prove this algorithm returns the desired result.



Proof. $|\psi_1\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$.

Now, since the oracle's action is dependent on the function, we do casework on the function.

Let the subscript c denote the function with $f(0) = f(1)$, and let the subscript b denote the function with $f(0) \neq f(1)$.

By noting that global phases don't matter, we get

$|\psi_2\rangle_c = |\psi_1\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$

$|\psi_2\rangle_b = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$

Thus

$|\psi_3\rangle_c = |0\rangle$

$|\psi_3\rangle_b = |1\rangle$

as desired.

■

6. Continued Fractions

Before we proceed with Shor's algorithm, we discuss the basics of continued fractions.

Definition 6.1. A **continued fraction** is a possibly infinite sequence of numbers $a_1, a_2, a_3, \ldots$ representing positive number r. Specifically, let $r_0 = r$. Then, a_n and r_n are defined by

$a_n = \lfloor r_n \rfloor$ (if r_n is an integer we terminate the sequence OR let $a_n = \lfloor r_n \rfloor - 1, a_{n+1} = 1$ and then terminate the sequence. $r_{n+1} = \frac{1}{r_n - a_n}$.

Then, we have

$r = a_0 + \frac{1}{a_1 + \frac{1}{1 + \frac{1}{a_2 + \ldots}}}$.

Definition 6.2. Let the nth convergent of the continued fraction of p be the continued fraction of p truncated to n terms.

Definition 6.3 (Continuants)**.** Let $a_1, \ldots, a_n$ be a sequence of integers. The **continuant** $K(a_1, \ldots, a_n)$ is defined recursively by

$$K() = 1, \qquad K(a_1) = a_1,$$

and for $n \geq 2$,

$$K(a_1, \ldots, a_n) = a_n K(a_1, \ldots, a_{n-1}) + K(a_1, \ldots, a_{n-2}).$$

This definition may seem a bit uncalled for, but it is because we can express a continued fraction in terms of the ratio of continuants.

Theorem 6.4 (Continuants and continued fractions)**.** *Let*

$$[a_1; a_2, \ldots, a_n] = a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}$$

be a finite continued fraction. Then

$$[a_1; a_2, \ldots, a_n] = \frac{K(a_1, \ldots, a_n)}{K(a_2, \ldots, a_n)}.$$

Note that these fractions are in their simplest form.

Theorem 6.5. *Let the nth convergent be written as $\frac{p_n}{q_n}$ (in most simplified form). Then $p_n = a_n p_{n-1} + p_{n-2}, q_n = a_n q_{n-1} + q_{n-2}$.*

Proof. $p_n = K(a_1, \ldots, a_n) = a_n K(a_1, \ldots, a_{n-1}) + K(a_1, \ldots, a_{n-2}) = a_n p_{n-1} + p_{n-2}$.

$q_n = K(a_2, \ldots, a_n) = a_n K(a_2, \ldots, a_{n-1}) + K(a_2, \ldots, a_{n-2}) = a_n q_{n-1} + q_{n-2}$ as desired.

Note that $p_{-2} = 0$, $p_{-1} = 1$, $q_{-2} = 1$, $q_{-1} = 0$ works as a negative index (because $p_0 = a_n$, $q_0 = 1$ by plugging in the reucurrence, which is correct).

■

Theorem 6.6. *Let the nth convergent be written as $\frac{p_n}{q_n}$ (in most simplified form). Then $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$.*

Proof. We have

$$\begin{pmatrix} 0 & 1 \\ 1 & a_{n+1} \end{pmatrix} \begin{pmatrix} p_{n-1} & q_{n-1} \\ p_n & q_n \end{pmatrix} = \begin{pmatrix} p_n & q_n \\ p_{n+1} & q_{n+1} \end{pmatrix}.$$

Hence, by multiplicity of determinants, we have $-(p_n q_{n-1} - p_{n-1} q_n = (-1)^n)$, so $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$.

■

Theorem 6.7 (Legendre's Theorem)**.** *Let $\frac{p}{q}$ be a fraction. If $|x - \frac{p}{q}| \leq \frac{1}{2q^2}$, then $\frac{p}{q}$ (in simplified form) is a convergent of x.*

Proof. We follow the proof given in [HW79].

Let us write $x - \frac{p}{q} = \frac{\theta}{q^2}$, where $0 < \theta < \frac{1}{2}$.

Let the convergents of $\frac{p}{q}$ be $\frac{p_0}{q_0}, \frac{p_1}{q_1}, \ldots, \frac{p_n}{q_n}$, where n is odd (we can choose this because of the final ambiguity stated in the definition of continued fractions).

Let $\omega = \frac{1}{\theta} - \frac{q_{n-1}}{q_n}$, or $\theta = \frac{q_n}{q_{n-1} + \omega q_n}$. Note that $\omega > 1$.

Thus

$$\begin{aligned} x &= \frac{p}{q} + \frac{\theta}{q^2} \\ &= \frac{p_n}{q_n} + \frac{\theta}{q_n^2} \\ &= \frac{p_n}{q_n} + \frac{1}{q_n(q_{n-1} + \omega * q_n)} \\ &= \frac{(p_n q_n + (1)^{n-1}) + \omega p_n q_n}{q_n(q_{n-1} + \omega q_n)} \\ &= \frac{p_{n-1} + \omega p_n}{q_{n-1} + \omega q_n}. \end{aligned}$$

Let the continued fraction list of ω be k. We can write $x = [a_0; a_1, a_2, a_3, \dots a_n, k]$. Thus $\frac{p}{q}$ is a convergent of x, as desired.

■

7. Shor's Algorithm

We now examine Shor's algorithm for factoring $N = pq$. We will assume in this algorithm that N is odd, because this can be easily checked. The algorithm uses both classical and quantum computers for factoring. The procedure of this algorithm is as follows.

(1) Pick a number a such that $1 < a < N$. If $\gcd(a, N) = 1$, proceed. Otherwise, the gcd is a prime factor of N, and we are done.
(2) Find the order of a for modulo N. Let the order be r.
(3) If r is odd or $a^{\frac{r}{2}} = -1 \pmod N$ then go back to step 1 by picking a different a.
(4) Otherwise, we know $(a^{\frac{r}{2}} - 1)(a^{\frac{r}{2}} + 1) = 0 \pmod N$ and none of the two numbers are divisible by N. Hence, compute $\gcd(a^{\frac{r}{2}} + 1, N)$ and you will get one of the prime factors.

Now, we state two theorems.

Theorem 7.1. *Let p be a prime, and let k be a natural number. If $k \mid p-1$, then the number of integers $a \in [1, p-1]$ such that the order of a modulo p is k is exactly $\varphi(k)$.*

Proof. It is clear that $k|p-1$ is required.

Let y be a primitive root in modulo p. Then, let $x = y^a$. We therefore have that the order of x modulo p is $\frac{p-1}{\gcd(p-1,a)}$. For this to equal k, we need a to be a multiple of $\frac{p-1}{k}$ and relatively prime to k, while not exceeding k (since $a < p$). This is exactly $\varphi(k)$, as desired.

■

Theorem 7.2. *Let $N = pq$ for primes p and q that are not 2. Then*
$\mathbb{P}(r$ *is odd or* $a^{\frac{r}{2}} \neq -1 \pmod N) \leq \frac{1}{2}$ *for a randomly chosen a between 1 and N.*

Proof. $\mathbb{P}(r$ is odd or $a^{\frac{r}{2}} = -1 \pmod N) \leq \mathbb{P}(r = 1 \pmod 2) + \mathbb{P}(a^{\frac{r}{2}} = -1(\pmod N))$.

Consider modulo p and modulo q separately.

We know the order of a modulo N is the least common multiple of the order of a modulo p and the order of a modulo q by the Chinese Remainder Theorem. The probability of the order a modulo p is odd, by our formula from earlier, is
$\frac{\sum_{i|(p-1), i=1 \pmod 2} \varphi(i)}{p-1} = \frac{1}{v_2(p-1)} \geq \frac{1}{2}$.

This is from the identity $\sum_{i|k} \varphi(i) = k$.

Hence, the probability of the order a modulo N is odd is at most $\frac{1}{4}$.

Now, for the second term, we also consider modulo p and modulo q. By Chinese Remainder Theorem, $a^{\frac{r}{2}} = -1 \pmod{p}$ and $a^{\frac{r}{2}} = -1 \pmod{q}$.

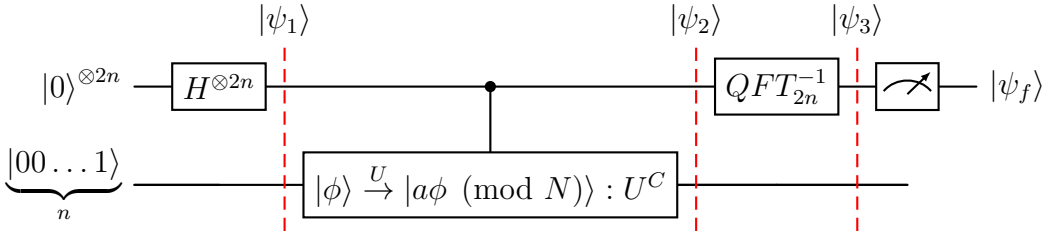
This means that 2-adic valuation of the order of a modulo p is the same as the 2-adic valuation of the order of a modulo q, and they are both even. Now, the probability that the 2-adic valuation of the order of a modulo p is x is $\frac{\sum_{i|(p-1),v_2(i)=x} \varphi(i)}{p-1} = \frac{1}{v_2(p)} \leq \frac{1}{2}$. Hence, the probability the 2-adic valuation of the order of a modulo p is the same as the 2-adic valuation of the order of a modulo q is at most $\frac{1}{4}$.

$\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, so we are done. ■

Hence, in the third step of Shor's algorithm we will proceed with a probability of at least $\frac{1}{2}$, and most likely higher. Notice that calculating the gcd is extremely efficient using Euclid's algorithm (time complexity is at most $O(n^2)$, where n is the number of bits because we have at most $O(n)$ subtractions and each subtraction takes at most $O(n)$ time because there are at most n digits).

Thus, the only non-polynomial part of this algorithm for a classical computer is step 2. Thus, we use a quantum computer for this part.

The quantum algorithm circuit diagram for finding the order is below. Let r be the period. Let N be written in binary with n bits.



Remark 7.3. The unitary operator in the circuit diagram above acts as identity on $N < \phi < 2^N$. This is not important to the algorithm but needs to be included so U is a unitary operator.

Also, for Shor's algorithm to be physically implemented, there needs to be 3 registers. This is because when one converts from classical gates to quantum gates, an ancillary is required. However, in this paper, we have neglected the ancillary.

We will write the current quantum state in two forms, one assuming knowledge of the period, and one assuming otherwise. They are denoted as n (no knowledge) and p (know the period).

$$|\psi_1\rangle_n = \frac{1}{2^n} \sum_{j=0}^{2^{2n}-1} |j\rangle\,|1\rangle$$

$$|\psi_1\rangle_p = \frac{1}{2^n\sqrt{r}} \sum_{j=0}^{2^{2n}-1} \sum_{x=0}^{r-1} \frac{1}{\sqrt{r}} \sum_{y=0}^{r-1} e^{\frac{-2\pi i x y}{r}} |j\rangle\,|a^y \pmod{N}\rangle$$

This is because
$\frac{1}{r} \sum_{x=0}^{r-1} \sum_{y=0}^{r-1} e^{\frac{-2\pi i x y}{r}} |a^y\rangle = \frac{1}{r} \sum_{y=0}^{r-1} r\delta_{0,y} |a^y\rangle = |1\rangle$ as desired.

$$|\psi_2\rangle_n = \frac{1}{2^n}\sum_{j=0}^{2^{2n}-1}|j\rangle\,|a^j \pmod N\rangle$$

.

$$|\psi_2\rangle_p = \frac{1}{2^n\sqrt{r}}\sum_{j=0}^{2^{2n}-1}\sum_{x=0}^{r-1}\frac{1}{\sqrt{r}}\sum_{y=0}^{r-1}e^{\frac{-2\pi ixy}{r}}|j\rangle\,|a^{y+j} \pmod N\rangle =$$

$$\frac{1}{2^n\sqrt{r}}\sum_{j=0}^{2^{2n}-1}\sum_{x=0}^{r-1}\frac{1}{\sqrt{r}}e^{\frac{2\pi ixj}{r}}\sum_{y=0}^{r-1}e^{\frac{-2\pi ixy}{r}}|j\rangle\,|a^{y} \pmod N\rangle$$

$$|\psi_3\rangle_n = \frac{1}{2^{2n}}\sum_{j=0}^{2^{2n}-1}\sum_{k=0}^{2^{2n}-1}e^{-\frac{2\pi ijk}{2^{2n}}}|k\rangle\,|a^j \pmod N\rangle\,.$$

$$|\psi_3\rangle_p = \frac{1}{2^{2n}\sqrt{r}}\sum_{j=0}^{2^{2n}-1}\sum_{k=0}^{2^{2n}-1}\sum_{x=0}^{r-1}\frac{1}{\sqrt{r}}e^{2\pi ij(\frac{2^{2n}x-rk}{2^{2n}r})}\sum_{y=0}^{r-1}e^{\frac{-2\pi ixy}{r}}|k\rangle\,|a^y \pmod N\rangle\,.$$

Now, for $|\psi_f\rangle$, a measurement is performed on the first register. We rewrite $|\psi_3\rangle_p$ into a cleaner form, in which $|\alpha_x\rangle = \sum_{y=0}^{r-1}e^{\frac{-2\pi ixy}{r}}|a^y \pmod N\rangle$. Note that $|\alpha_0\rangle, |\alpha_1\rangle, \ldots, |\alpha_{n-1}\rangle$ form an orthonormal unit basis.

$$|\psi_3\rangle_p = \frac{1}{2^{2n}\sqrt{r}}\sum_{x=0}^{r-1}\sum_{j=0}^{2^{2n}-1}\sum_{k=0}^{2^{2n}-1}e^{2\pi ij(\frac{2^{2n}x-rk}{2^{2n}r})}|k\rangle\,|\alpha_x\rangle$$

.

Thus, upon measurement of the first register (abbreviated $|\phi\rangle$), we have

$\mathbb{P}(|\phi\rangle = |k\rangle) = \frac{1}{2^{4n}r}\sum_{x=0}^{r-1}|\sum_{j=0}^{2^{2n}-1}e^{2\pi ij(\frac{2^{2n}x-rk}{2^{2n}r})}|^2 = \frac{1}{2^{4n}r}\sum_{x=0}^{r-1}|\frac{1-e^{2\pi i\frac{2^{2n}x-rk}{r}}}{1-e^{2\pi i(\frac{2^{2n}x-rk}{2^{2n}r})}}|^2$

To simplify this further, we recall an elementary identity from precalculus:

$|1-e^{i\phi}|^2 = (1-e^{i\phi})(1-e^{-i\phi}) = 2-e^{i\phi}-e^{-i\phi} = 2-2\cos(\phi) = 4\sin(\frac{\phi}{2})^2$.

Abbreviating $\delta_x = \frac{(2^{2n}x-rk))}{2^{2n}r}$ (do not confuse this with $\delta_{i,j}$), we have

$\mathbb{P}(|\phi\rangle = |k\rangle) = \frac{1}{2^{4n}r}\sum_{x=0}^{r-1}|\frac{\sin(\pi 2^{2n}\delta_x)}{\sin(\pi\delta_x)}|^2$

This is very clean.

Now, we will make several approximations. Note that Shor's algorithm is in complexity class **BQP**, meaning it successfully runs with $O(1) \geq \frac{2}{3}$ probability of success. However, if $O(1) < \frac{2}{3}$, then we can repeat the algorithm $O(1)$ times to get our desired success rate, and thus it won't change quantum gate complexity. Thus, we prove that Shor's algorithm runs successfully $O(1)$ percent of the time.

We notice that there is an appropriate choice of k such that for one of the x (specifically, the closest approximation of $\frac{2^{2n}x}{r}$ as an integer), $|\delta_x| \leq \frac{1}{2^{2n+1}}$. Call that value of k such that this occurs o_x. From Taylor expansion (there are more accurate methods like the slope of the secant line from $x=0$ to $x=\frac{\pi}{2}$, but this will suffice), we have

$\mathbb{P}(|\phi\rangle = |o_x\rangle) \geq \frac{1}{2^{4n}r}|\frac{\sin(\pi 2^{2n}\delta_x)}{\sin(\pi\delta_x)}|^2 \geq \frac{1}{2^{4n}r}|\frac{\pi 2^{2n}\delta_x - \frac{\pi^3 2^{6n}\delta_x^3}{6}}{\pi\delta_x}|^2 \geq \frac{1}{2^{4n}r}|2^{2n}(1-\frac{\pi^2}{24})|^2 \geq \frac{1}{r}(1-\frac{\pi^2}{24})^2$.

Note that $\frac{\pi^2}{24} \approx 0.35$.

Therefore, with probability at least 35%, our measurement returns the closest approximation of $2^{2n}\frac{x}{r}$ as an integer, where x is a random number between 0 and r that we do not know. Because of Euler's Generalization of Fermat's Little Theorem, we have $r < N < 2^n$.

Now, let $A = \frac{x}{r}$, and let $B = \frac{\lfloor \frac{2^{2n}x}{r} \rceil}{2^{2n}}$. We know B, and we wish to find A.

We know $|A - B| \leq \frac{1}{2^{2n+1}} < \frac{1}{2N^2} < \frac{1}{2r^2}$ and the denominator of A is less than N, by Legendre's theorem, A is a convergent of B. Now, we claim that for fraction S with denominator less than N such that $|S - B| < \frac{1}{2N^2}$ implies $S = A$. Otherwise, $|S - A| > \frac{1}{N^2}$, impossible.

Thus, we search all the convergents up to denominator N and take the first one such that the inequality from Legendre's holds. Therefore, we will be able to retrieve $\frac{x}{r}$.

When we retrieve $\frac{x}{r}$, if x is not relatively prime to r, our denominator will not be the actual value of r. This can be resolved. Run the algorithm twice. Then, we will get numbers $\frac{x_1}{r}, \frac{x_2}{r}$. The chance that the least common multiple of the two denominators is not r is at most (a summation of all primes) $\Pi_p(1 - \frac{1}{p^2}) \leq \Pi_{n=2}^{\infty}(1 - \frac{1}{n^2}) = \Pi_{n=2}^{\infty}\frac{(n-1)(n+1)}{n^2} = \frac{1}{2}$ which is $O(1)$.

Thus, the error in Shor's algorithm is bounded by $O(1)$.

7.1. **Time Complexity.** We now examine the time complexity of Shor's algorithm. The classical parts of Shor's algorithm are very fast, so we ignore them. Instead, notice that the Hadamard gate takes $2n$ basic quantum gates, and the QFT inverse takes at most $O(n^2)$ quantum gates. Now, for the controlled unitary, we realize that the classical time of this gate is greater than or equal to the quantum time of this gate because the Toffoli gates are complete, so we can convert classical gates to quantum gates.

Now, our classical function has inputs C, ϕ and outputs $a^C * \phi \pmod N$. We can create a search table for each C a power of 2 (we store around 2048 values). Thus, the number of multiplications required is the number of ones in C plus 1, which is $O(n)$ multiplications. The standard method of multiplication takes $O(n^2)$ time. Thus, the quantum time complexity of Shor's algorithm is $O(n^3)$ which is polynomial in the number of bits, n. This is much faster than the general number field sieve, which is sub-exponential, demonstrating the power of Shor's algorithm.

Acknowledgments

The author thanks Simon Rubinstein-Salzedo and Humza Hussaini for helpful suggestions on improving this paper.

References

[BLP93] J. P. Buhler, H. W. jun. Lenstra, and Carl Pomerance. Factoring integers with the number field sieve. In *The development of the number field sieve*, pages 50–94. Berlin: Springer-Verlag, 1993.

[Gle57] Andrew M. Gleason. Measures on the closed subspaces of a Hilbert space. *J. Math. Mech.*, 6:885–893, 1957.

[HW79] G. H. Hardy and E. M. Wright. An introduction to the theory of numbers. 5th ed. Oxford etc.: Oxford at the Clarendon Press. xvi, 426 p. hbk: £ 17.50; pbk: £ 8.00 (1979)., 1979.

[Sho94] P.W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, 1994.