SHOR'S ALGORITHM

MANU ISAACS

1. Abstract

In this paper, we discuss Shor's algorithm, one of the most important algorithms in quantum computing. Shor's algorithm can factor the M in polynomial time (in $\log M$). Even the best classical factoring algorithms run in superpolynomial time. Because many cryptographic algorithms rely on the difficulty of factoring, Shor's algorithm provides the means for quantum computers to undermine the security of the internet. Shor's algorithm begins by turning the problem of factoring into a problem of finding the period of a function. Then, using quantum parallelism on a superposition of inputs, the function is computed over many values at once. Then, to find the period of this function, we take the quantum Fourier transform of the resulting quantum state. Finally, we measure the system, obtaining a number which, with a high probability, is a nontrivial factor of M.

2. Introduction

Shor's algorithm was discovered in 1994 by mathematician Peter Shor. The algorithm showed the promise of quantum computers to undermine the security of much of cryptography. However, actual quantum computers of the required size have yet to be built - the largest number factored on a quantum computer using Shor's algorithm is only 21.

This paper starts by introducing the basics of quantum computation, assuming no background. We begin by defining a single qubit and single qubit measurement. Then, we progress to multi-qubit systems and multi-qubit measurement. Then, we discuss quantum state transformations, and name some of the commonly used quantum gates (or transformations). Then, we look at how we can use these gates to build up the quantum analog of classical computers.

Then, we move on to the quantum Fourier transform, which is essential to Shor's algorithm. It is very similar to the discrete Fourier transform. Finally we move on to Shor's algorithm itself. We describe how finding a factor of M can be reduced to finding the period of the function $f(k) \equiv a^k \pmod{M}$ using elementary number theory. Then, we describe how to perform the quantum part of the algorithm, which involves applying f and a quantum Fourier transform to a quantum superposition. We show that after measurement, we will get a nontrivial factor of M with high probability. We end by looking at generalizations of Shor's algorithm and some of the open problems in quantum computation. as well as the cryptographic implications of Shor's Algorithm.

3. Acknowledgements

The author would like to thank the Euler Circle, Sawyer Dobson, and Simon Rubenstein-Salzedo for making this paper possible.

Date: July 17, 2023.

4. What is a qubit?

4.1. **Linear Algebra Review.** At the beginning of some sections, a review of the necessary linear algebra is given, assuming the reader has no prior experience with linear algebra. Feel free to skip these sections if you don't think you need them. However, in this section, Dirac's braket notation, which is used in quantum computing and throughout this paper is introduced, so unless you are familiar with it, it is recommended that you at least skim this section.

Definition 4.1. A *vector space* V is a set of vectors which can by multiplied by scalars in a field F and added to one another. Throughout this paper, we will be working in complex vector spaces, that is to say the $F = \mathbb{C}$, and we will denote vectors using Dirac's braket notation $|v\rangle$. $|v\rangle$ is called a *ket*. Vector spaces satisfy the following axioms for $1, a, b \in F$ and $|v\rangle, |u\rangle, |w\rangle \in V$:

- $(|v\rangle + |u\rangle) + |w\rangle = |v\rangle + (|u\rangle + |w\rangle)$ (Associativity of vectors)
- $|v\rangle + |w\rangle = |w\rangle + |v\rangle$ (Commutativity of vectors)
- $\exists\, |0\rangle \in V : |0\rangle + |v\rangle = |v\rangle \ \forall v \in V$ (Existence of zero vector)
- $\forall\, |v\rangle \in V, \exists\, |-v\rangle \in V : |v\rangle + |-v\rangle = |0\rangle$ (Existence of the additive inverse)
- $a(b\,|v\rangle) = (ab)\,|v\rangle$ (Associativity of scalar multiplication)
- $1\,|v\rangle = |v\rangle$ (Existence of scalar multiplicative identity)
- $a(|v\rangle + |u\rangle) = a\,|v\rangle + a\,|u\rangle$ (Vector distributive property of scalar multiplication)
- $(a + b)\,|v\rangle = a\,|v\rangle + b\,|v\rangle$ (Scalar distributive property of scalar multiplication)

Definition 4.2. Given a set $B = \{|\beta_1\rangle, |\beta_2\rangle, \ldots |\beta_n\rangle\}$ of vectors in V, we say that a vector $|w\rangle$ is a *linear combination* of B if

$$B = \sum_{i=1}^{n} a_i\,|\beta_i\rangle$$

for some scalar a_i.

Definition 4.3. Given a set S of vectors in V, the *span of* S, denoted $\text{span}(S)$, is the set of vectors in V which are linear combinations of S. Conversely, $\text{span}(S)$ is said to be *generated* by S.

Definition 4.4. Given $|v\rangle, |w\rangle \in V$, let the *inner product* of $|v\rangle$ and $|w\rangle$, denoted by $\langle v|w\rangle$, satisfy

$$\langle v|v\rangle \in \mathbb{R}^+$$
$$\langle v|w\rangle = \overline{\langle w|v\rangle}$$
$$\langle v|a\,|w\rangle + b\,|u\rangle\rangle = a\,\langle v|w\rangle + b\,\langle v|u\rangle$$

Definition 4.5. A set of vectors $|v_1\rangle, |v_2\rangle, \ldots, |v_n\rangle$ in a vector space V is said to be *orthogonal* if every pair of distinct vectors in the set is orthogonal, i.e., $\langle v_i|v_j\rangle = 0$ for all $i \neq j$.

Definition 4.6. The *norm* of a vector $|v\rangle$ in a vector space V, denoted by $||v\rangle|$, represents the length or magnitude of the vector. It is defined by $\sqrt{\langle v|v\rangle}$.

Definition 4.7. A set of vectors $|v_1\rangle, |v_2\rangle, \ldots, |v_n\rangle$ in a vector space V is said to be *orthonormal* if it is orthogonal and every vector in the set has unit norm, i.e., $\langle v_i|v_j\rangle = \delta_{ij}$, where δ_{ij} is the Kronecker delta symbol, defined as

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Definition 4.8. A *basis* for a vector space V is a set of linearly independent vectors that span the entire space. In other words, a basis is a set of vectors $B = |\beta_1\rangle, |\beta_2\rangle, \ldots, |\beta_n\rangle$ such that any vector $|v\rangle$ in V can be expressed as a unique linear combination of the basis vectors:

$$|v\rangle = \sum_{i=0}^{n} c_i |\beta_i\rangle$$

where c_i are scalars. Note that $\text{span}(B) = V$.

Definition 4.9. An *orthonormal basis* is a basis for a vector space V that is also an orthonormal set. In this paper, any basis is assumed to be an orthonormal basis.

Definition 4.10. For a given basis $B = \{|\beta_1\rangle, |\beta_2\rangle, \ldots, |\beta_n\rangle\}$ of vector space V, if some $|v\rangle \in V$ can be written as

$$|v\rangle = \sum_{i=0}^{n} a_i |\beta_i\rangle$$

then in *vector notation*, we can write $|v\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$.

Definition 4.11. For two vectors $|v\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ and $|w\rangle = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$, define the dot product of $|v\rangle$ and $|w\rangle$ as

$$|v\rangle \cdot |w\rangle = \sum_{i=0}^{n} a_i b_i.$$

Definition 4.12. *Matrix multiplication* is an operation defined for two matrices A and B such that the number of columns in A is equal to the number of rows in B. The product of matrices A and B, denoted as AB, is a matrix whose elements are computed as follows:

If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then the product AB is an $m \times p$ matrix defined by:

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

where a_{ik} denotes the element in the i-th row and k-th column of A, and b_{kj} denotes the element in the k-th row and j-th column of B. We can also think of $(AB)_{ij}$ as the dot product of the i-th row of A and the j-th column of B.

Example. For example, consider the following matrix multiplication. Let

$$A = \begin{pmatrix} 2 & 3 \\ -1 & 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 5 & 1 \\ 2 & -3 \end{pmatrix}$$

To compute the product AB, we will calculate each element of the resulting matrix:

$$AB = \begin{pmatrix} (AB)_{11} & (AB)_{12} \\ (AB)_{21} & (AB)_{22} \end{pmatrix}$$

Let's perform the calculations:

$$\begin{aligned} (AB)_{11} &= (2 \cdot 5) + (3 \cdot 2) = 10 + 6 = 16 \\ (AB)_{12} &= (2 \cdot 1) + (3 \cdot -3) = 2 - 9 = -7 \\ (AB)_{21} &= (-1 \cdot 5) + (4 \cdot 2) = -5 + 8 = 3 \\ (AB)_{22} &= (-1 \cdot 1) + (4 \cdot -3) = -1 - 12 = -13 \end{aligned}$$

Therefore, the matrix product AB is:

$$AB = \begin{pmatrix} 16 & -7 \\ 3 & -13 \end{pmatrix}$$

So, the resulting matrix AB is a 2×2 matrix with the entries shown above.

Definition 4.13. Given a matrix A with entries a_{ij}, we define the *transpose conjugate* of A, denoted $B = A^\dagger$ as the matrix with entries $b_{ij} = \overline{a_{ji}}$

Example. If

$$A = \begin{pmatrix} 1 & 2 \\ 3i & 4i \end{pmatrix}$$

then

$$A^\dagger = \begin{pmatrix} 1 & -3i \\ 2 & -4i \end{pmatrix}.$$

Definition 4.14. Given a ket $|v\rangle$, we define its corresponding *bra* as $\langle v| = |v\rangle^\dagger$. So if

$$|v\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix},$$

then

$$\langle v| = |v\rangle^\dagger = \begin{pmatrix}\overline{a_1} & \overline{a_2} & \dots & \overline{a_n}\end{pmatrix}$$

Definition 4.15. We define the *standard inner product* of $|v\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ and $|w\rangle = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$, to be

$$\langle v|w\rangle = \langle v|\,|w\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix}\overline{b_1} & \overline{b_2} & \dots & \overline{b_n}\end{pmatrix} = \sum_{i=0}^{n} a_i\overline{b_i}$$

Note that this satisfies all the criteria of 4.4. From here onwards, when we refer to the inner product, we mean the standard inner product.

Example. Consider the vectors $|v\rangle = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $|w\rangle = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ which are perpendicular in the plane. We have

$$\langle v|w\rangle = \langle v|\,|w\rangle = \begin{pmatrix} 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = (2)(1) + (1)(-2) = 0,$$

so the vectors are orthogonal, as expected.

4.2. **Definition of a Qubit.**

Definition 4.16. A *qubit*, short for quantum bit, is the smallest unit of quantum information. For the purposes of this paper, a qubit can be thought of as a two-element vector, where each element is a complex (or real) number and the sum of the squares of the magnitudes of the elements of the vector is 1.

4.3. **Braket Notation For Qubits.** We define $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Notice that $|0\rangle$ is **not** the zero vector. The basis $\{|0\rangle, |1\rangle\}$ generates the vector space of a single qubit and is called the *standard basis*. Unless otherwise specified, this is the basis that we are working in. Then, we can write any qubit $|\psi\rangle$ as

$$|\psi\rangle = \begin{pmatrix} a \\ b \end{pmatrix} = a\,|0\rangle + b\,|1\rangle,$$

where $a, b \in \mathbb{C}$ and $|a|^2 + |b|^2 = 1$. The notation $|\psi\rangle$ is called a *ket* and represents the state of one or multiple qubits.

4.4. **Relative and global phases.** There is some redundancy to the above definition when considering the physical representation of a qubit.

Definition 4.17. If, for qubits $|v\rangle$ and $|w\rangle$, we can write $|v\rangle = c\,|w\rangle$ for some $c \in \mathbb{C}$, we say that $|v\rangle$ and $|w\rangle$ differ by a *global phase shift* of c. Furthermore, $|v\rangle$ and $|w\rangle$ have the same physical representation.

Definition 4.18. We define the *relative phase* of a qubit $|\phi\rangle = a|0\rangle + b|1\rangle$ to be the complex number c such that

$$\frac{a}{b} = c\frac{|a|}{|b|}.$$

Notice that if two qubits have different relative phases then they have different physical representations, and if two qubits have the same relative phase, then they have the same physical representation.

4.5. **Commonly used qubits.** The following qubits are use commonly enough to have their own notation:

$$\begin{aligned}
|+\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \\
|-\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle), \\
|i\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle), \\
|-i\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle).
\end{aligned}$$

Notice that because all of these qubits have different relative phases, they all have different physical representations. The basis $\{|+\rangle, |-|\rangle\}$ is called the Hadamard basis.

Example. Simplify $\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$.

We have

$$\begin{aligned}
\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle) &= \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) + \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\right) \\
&= \frac{1}{2}(|0\rangle + |1\rangle + |0\rangle - |1\rangle) \\
&= |0\rangle.
\end{aligned}$$

5. Measurement of a single qubit

The first operation on a qubit that we will look at is measurement. This is the only way in which we can physically observe quantum states. Interestingly, in observing a quantum state, we also change it to the state that we measured. In the following sections, we generalize this idea to multiple qubit systems.

Definition 5.1. By *measuring* the qubit $|\phi\rangle = a|\beta_1\rangle + b|\beta_2\rangle$ in the basis $\{\beta_1, \beta_2\}$, we observe the state as the following:

- $|\beta_1\rangle$ with probability $|a|^2$
- $|\beta_2\rangle$ with probability $|b|^2$

Furthermore, when we measure a qubit, the state of the qubit also changes to the state that it is measured as.

Note that performing the same measurement twice has the exact same effect as performing that measurement once. The order in which measurements are performed does matter, and performing measurement A followed by measurement B is not the same as performing measurement B.

Example. What happens when we measure the qubit $|0\rangle$ first in the Hadamard basis $\{|+\rangle, |-\rangle\}$ and then in the standard basis? To compute the probabilities for measurement in the standard basis, we have to write $|0\rangle$ in the Hadamard basis: $|0\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$. In this case, we see that $|a|^2 = |b|^2 = \frac{1}{2}$, so after measuring in the Hadamard basis, the qubit becomes either $|+\rangle$ or $|-\rangle$ each with equal probability. Then, recalling the definitions of $|+\rangle$ and $|-\rangle$, we see that measurement of either of them in the standard basis results in $|0\rangle$ or $|1\rangle$ again with equal chance.

Definition 5.2. We say that $|\phi\rangle = a|\beta_1\rangle + b|\beta_2\rangle$ is a *superposition* of the basis $\{\beta_1, \beta_2\}$ if a and b are both nonzero. If no basis is given, it is assumed that we are talking about the standard basis.

Remark 5.3. After measuring some qubit in the basis B, that qubit will not be in a superposition of the basis B.

Remark 5.4. Note that if two qubits $|v\rangle$ and $|w\rangle$ differ by a global phase, that is, they have the same representation, then the result of measuring $|v\rangle$ and $|w\rangle$ is the same no matter which basis we measure in. This is why we say $|v\rangle$ and $|w\rangle$ have the same physical representation.

6. Multiple qubit systems

6.1. Linear Algebra Review.

Definition 6.1. The *direct sum* $U = V \oplus W$ of vector spaces V with basis A and W with basis B where A and B are distinct is the vector space with basis $A \cup B$.

Remark 6.2. For all $|u\rangle \in U$ there exists unique $|v\rangle \in V$ and $|w\rangle \in W$ such that $|u\rangle = |v\rangle + |w\rangle$. This can be verified easily by looking at the representation of $|u\rangle, |v\rangle, |w\rangle$ in their respective bases.

Remark 6.3. The *dimension* of a vector space V, denoted by $\dim(V)$, is the cardinality or size of its basis B, $|B|$. Notice that $\dim(V \oplus W) = \dim(V) + \dim(W)$. In this sense, the dimension of vector spaces grows linearly with the direct sum.

Example. Suppose $V_1 = \mathbb{R}^2$ describes the state of a particle in a plane. An element $v_1 \in V_1$ could be written as $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = x_1|0_1\rangle + y_1|1_1\rangle$, where the basis of V_1 is $\{|0_1\rangle, |1_1\rangle\}$. Let V_2 be defined similarly. Then, $V_1 \oplus V_2$ describes all possible states of both particles in the plane. An element of $V_1 \oplus V_2$ could be written as

$$\begin{pmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{pmatrix} = x_1|0_1\rangle + y_1|1_1\rangle + x_1|0_2\rangle + y_1|1_2\rangle.$$

Definition 6.4. The tensor product $U = V \otimes W$ of vector spaces V with basis $A = \{|\alpha_1\rangle, |\alpha_2\rangle, \ldots |\alpha_n\rangle\}$ and W with basis $B = \{|\beta_1\rangle, |\beta_2\rangle, \ldots |\beta_n\rangle\}$ is the vector space with basis consisting of the nm elements of the form $|\alpha_i\rangle \otimes |\beta_j\rangle$. The tensor product on vectors satisfies the following properties:

(1) $(|v_1\rangle + |v_2\rangle) \otimes |w\rangle = |v_1\rangle \otimes |w\rangle + |v_2\rangle \otimes |w\rangle$ (Left distributive property)

(2) $|w\rangle \otimes (|v_1\rangle + |v_2\rangle) = |w\rangle \otimes |v_1\rangle + |w\rangle \otimes |v_2\rangle$ (Right distributive property)

(3) $(a\,|v\rangle) \otimes |w\rangle = a(|v\rangle \otimes |w\rangle) = |v\rangle \otimes (a\,|w\rangle)$ (Linearity)

Remark 6.5. Note that $\dim(V \otimes W) = \dim(V)\dim(W)$.

Remark 6.6. For the rest of this paper, we will refer to the tensor product of vectors not as $|v\rangle \otimes |w\rangle$ but simply as $|v\rangle\,|w\rangle$.

Example. Let V have basis $\{|0_v\rangle, |1_v\rangle\}$ and W have basis $\{|0_w\rangle, |1_w\rangle\}$. Then, if $|v\rangle = \frac{1}{\sqrt{2}}(|0_v\rangle + |1_v\rangle)$ and $|w\rangle = \frac{1}{\sqrt{13}}(2\,|0_w\rangle + 3\,|1_w\rangle)$, we have

$$\begin{aligned} |v\rangle\,|w\rangle &= \frac{1}{\sqrt{2}}(|0_v\rangle + |1_v\rangle)\frac{1}{\sqrt{13}}(2\,|0_w\rangle + 3\,|1_w\rangle) \\ &= \frac{1}{\sqrt{26}}(2\,|0_v\rangle\,|0_w\rangle + 3\,|0_v\rangle\,|1_w\rangle + 2\,|1_v\rangle\,|0_w\rangle + 3\,|1_v\rangle\,|1_w\rangle) \end{aligned}$$

by the properties of the tensor product of vectors in 6.4.

6.2. **The Vector Space of a Multi-Qubit System.**

Definition 6.7. Consider now a system of n qubits, labeled 0 to $n-1$. Qubit i is in vector space V_i which has basis $\{|0_i\rangle, |1_i\rangle\}$. Then, the vector space of the entire n-qubit system is $V = V_0 \otimes V_1 \otimes \cdots \otimes V_n$.

Notice that by 6.5, the dimension of V is $N = 2^n$. This exponential increase in dimension indicates the power of quantum computer. We can write the basis of V as

$$\begin{gathered} \{|0\rangle_{n-1} \otimes |0\rangle_{n-2} \otimes \cdots \otimes |0\rangle_1 \otimes |0\rangle_0, \\ |0\rangle_{n-1} \otimes |0\rangle_{n-2} \otimes \cdots \otimes |0\rangle_1 \otimes |1\rangle_0, \\ |0\rangle_{n-1} \otimes |0\rangle_{n-2} \otimes \cdots \otimes |1\rangle_1 \otimes |0\rangle_0, \\ \vdots \\ |1\rangle_{n-1} \otimes |1\rangle_{n-2} \otimes \cdots \otimes |1\rangle_1 \otimes |1\rangle_0\} \end{gathered}$$

but this is extremely clunky. Let's remove the tensor products and subscripts, and combine the kets together:

$$\{|00\ldots00\rangle, |00\ldots01\rangle, |00\ldots10\rangle \ldots |11\ldots11\rangle\}$$

That's much better. A lot of time, we will use this notation to describe the basis of an n-qubit system. However, a further simplification is possible by writing every number in decimal as opposed to binary:

$$\{|0\rangle, |1\rangle, |2\rangle \ldots |2^n - 1\rangle\}.$$

We can also write the basis in vector notation, where each basis vector has N elements:

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\}.$$

Then, the standard basis for a two qubit system can be written as

$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$$

or

$$\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}.$$

or

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Example. Expand $|+\rangle |+\rangle$.

We have

$$\begin{aligned} |+\rangle |+\rangle &= \frac{1}{2}((|0\rangle + |1\rangle)(|0\rangle + |1\rangle)), \\ &= \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle), \\ &= \frac{1}{2}(|0\rangle + |1\rangle + |2\rangle + |3\rangle). \end{aligned}$$

We see a pattern here which generalizes to the following lemma, which turns out to be of paramount importance in Shor's Algorithm:

Lemma 6.8. *We claim that for positive integers n and $N = 2^n$,*

$$\underbrace{|+\rangle |+\rangle \dots |+\rangle}_{n \text{ times}} = \frac{1}{\sqrt{N}}(|0\rangle + |1\rangle + \dots |N-1\rangle).$$

Proof. We can rewrite the left hand side as $\frac{1}{\sqrt{N}} \underbrace{(|0\rangle + |1\rangle)(|0\rangle + |1\rangle) \dots (|0\rangle + |1\rangle)}_{n \text{ times}}$. For each term in the binomial expansion of this expression, we get a distinct number whose binary representation ranges from $|0\rangle$ to $|N-1\rangle$, but this is the right hand side, so we are done. ■

6.3. **Entangled and Unentangled States.**

Definition 6.9. An n-qubit system is in an *entangled state* if it cannot be written as the tensor product of single qubits, A state that is not entangled is *unentangled*.

Remark 6.10. The state $|\phi\rangle = \frac{1}{2}(|00\rangle - |01\rangle - |10\rangle + |11\rangle)$ is not entangled, but the state $|psi\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle - |11\rangle)$ is.

Proof. Notice that $|\phi\rangle = |-\rangle |-\rangle$ so $|\phi\rangle$ is unentangled. Now, assume contrary, that $|\psi\rangle$ is unentangled. Then, we can write

$$\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle - |11\rangle) = (a|0\rangle + b|1\rangle)(c|0\rangle + d|1\rangle),$$

so $ac = ad = bc = -bd = \frac{1}{4}$. But then we have $(ac)(bd) = abcd = \frac{-1}{16}$ while $(ad)(bc) = abcd = \frac{1}{16}$, contradiction. ■

7. Measurement of multiple qubits

7.1. **Definition of Multi-Qubit Measurement.** Any measuring device which measures an n-qubit state with vector space V is characterized by a set of vector spaces $\{S_i\}$ satisfying

$$V = S_1 \oplus S_2 \oplus \cdots \oplus S_k,$$

where $k \leq N$, where $N = 2^n$ is the dimension of V. What happens when we measure $|v\rangle$? We know that $|v\rangle$ can be written uniquely as

$$|v\rangle = \sum_{i=1}^{k} c_i \, |s_i\rangle$$

for unit vectors $|s_i\rangle \in S_i$. Then, when $|v\rangle$ is measured, we get the state $|s_i\rangle$ with probability $|c_i|^2$, and as before, the state itself changes to what we measure it as

8. Transformations

8.1. **Linear Review.**

Definition 8.1. We say a matrix U is *unitary* if $U^\dagger U = I$.

Theorem 8.2. *A matrix U is unitary if and only if it maps unit vectors to unit vectors.*

Proof. First, we show that if U is unitary, then it maps unit vectors to unit vectors. Let $|v\rangle$ be a unit vector (i.e., $\| \, |v\rangle \, \| = 1$) and let U be unitary. We want to show that $U \, |v\rangle$ is also a unit vector. Let's calculate the norm of $U \, |v\rangle$:

$$|U \, |v\rangle \, |^2 = \langle v|U^\dagger U|v\rangle \, \langle v|I|v\rangle = \langle v|v\rangle = \| \, |v\rangle \, \|^2 = 1.$$

This proves that U maps unit vectors to unit vectors when U is unitary. Now, we show that if U maps unit vectors to unit vectors, then it is unitary. Assume that U is a square matrix that maps unit vectors to unit vectors. We want to prove that U is unitary, i.e., $UU^\dagger = I$. Take an arbitrary unit vector $|x\rangle$. Since U maps unit vectors to unit vectors, we can say:

$$1 = |U \, |x\rangle \, | = \langle x|U^\dagger U|x\rangle \, .$$

Let $\{x_1, x_2, \ldots x_n\}$ be an orthonormal basis. Then, notice that $\langle x_i| \, UU^\dagger \, |x_j\rangle = (UU^\dagger)_{ij} = \delta_{ij}$ which means that $UU^\dagger = I$ and U is unitary, as desired. ∎

Corollary 8.3. *Unitary matrices map quantum states to quantum states.*

Theorem 8.4. *We claim that if U is unitary, then so is U^{-1}.*

Proof. Using the fact that $U^{-1} = U^\dagger$, we have

$$\begin{aligned} (U^{-1})^\dagger U^{-1} &= (U^\dagger)^\dagger U^{-1} \\ &= UU^{-1}, \\ &= I. \end{aligned}$$

so U^{-1} is unitary. ∎

8.2. **Unitary Transformations.** The other type of operation that we can perform on qubits are transformations. Transformations can be thought of as 2^n by 2^n unitary matrices which act on an n-qubit system. Notice that by 8.3 Transformations are also called gates.

8.3. **Basic Single-Qubit Gates.** For example, we have the following single-qubit transformations:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$
$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$Y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$
$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

The first four gates are the Pauli Transformations, and the last one is the Hadamard Transformation. Notice that the not gate sends $|0\rangle$ to $|1\rangle$ and vice versa:

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

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

Notice also that the Hadamard gate sends $|0\rangle$ to $|+\rangle$ and $|1\rangle$ to $|-\rangle$:

$$H|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = |+\rangle,$$

$$H|1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = |-\rangle.$$

8.4. **Basic Multi-Qubit Gates.**

8.4.1. *Controlled NOT (CNOT) Gate.* The Controlled NOT or CNOT gate is a two-qubit gate that applies the X gate to the target qubit if the control qubit is $|1\rangle$, and does nothing if the control qubit is $|0\rangle$. It leaves the control qubit unchanged. In the case where the target qubit is in a superposition, the result can be calculated using linearity or by using the below matrix. The matrix representation of the CNOT gate is:

$$\text{CNOT} = \bigwedge\nolimits_{1} X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

8.4.2. *Controlled Gates.* In addition to the CNOT gate, there are controlled versions of other single-qubit gates. These gates apply the single-qubit gate to the target qubit only when the control qubit is $|1\rangle$. For instance, the controlled-U gate (where U represents any single-qubit gate) applies the gate U to the target qubit when the control qubit is $|1\rangle$. The matrix representation of a controlled gate is similar to the CNOT gate, with the gate U appearing in the appropriate locations:

$$\bigwedge_1 U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & U_{00} & U_{01} \\ 0 & 0 & U_{10} & U_{11} \end{pmatrix}$$

8.4.3. *Toffoli Gate.* The Toffoli gate, also known as the Controlled-Controlled-NOT (CC-NOT) gate, is a three-qubit gate that applies the Pauli-X gate to the target qubit if both control qubits are $|1\rangle$. If either of the control qubits is $|0\rangle$, it leaves the target qubit unchanged. The matrix representation of the Toffoli gate is:

$$\text{Toffoli} = \bigwedge_2 X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

8.5. **A universally approximating set of gates.** The Solovey-Kiteav theorem says that there exists a finite gate set such that it is possible to construct any gate to accuracy 2^{-d} in $P(d)$ gates, where P is a polynomial [RP14]. One example of such a gate set is the set $\{X, T\}$. This result is not central to Shor's algorithm, but it is still important. We do not prove it in this paper, but you can read the proof in [DN05].

8.6. **The No-Cloning Principle.**

Theorem 8.5. *Given some state $|a\rangle|0\rangle$, it is impossible to "clone" the $|a\rangle$ state, that is, it is impossible to have a transform*

$$U: \ |a\rangle|0\rangle \to |a\rangle|a\rangle$$

for all $|a\rangle$.

Proof. For some qubits $|a\rangle$ and $|b\rangle$, let $|c\rangle = \frac{1}{\sqrt{2}}(|a\rangle + |b\rangle)$. By the linearity of U, we have

$$\begin{aligned} U(|c\rangle|0\rangle) &= U(\frac{1}{\sqrt{2}}(|a\rangle + |b\rangle)|0\rangle) \\ &= \frac{1}{\sqrt{2}}(U(|a\rangle|0\rangle) + U(|b\rangle|0\rangle)) \\ &= \frac{1}{\sqrt{2}}(|a\rangle|a\rangle + |b\rangle|b\rangle) \end{aligned}$$

but we also have

$$\begin{aligned} U(|c\rangle|0\rangle) &= |c\rangle|c\rangle \\ &= \frac{1}{\sqrt{2}}(|a\rangle + |b\rangle)\frac{1}{\sqrt{2}}(|a\rangle + |b\rangle) \\ &= \frac{1}{2}(|a\rangle|a\rangle + |a\rangle|b\rangle + |b\rangle|a\rangle + |b\rangle|b\rangle) \end{aligned}$$

but

$$\frac{1}{2}(|a\rangle|a\rangle + |a\rangle|b\rangle + |b\rangle|a\rangle + |b\rangle|b\rangle) \neq \frac{1}{\sqrt{2}}(|a\rangle|a\rangle + |b\rangle|b\rangle),$$

so it is impossible to "clone" a qubit $|a\rangle$. ■

9. Quantum Analog of Classical Gates

9.1. **Reversibility.** All quantum state transformations are reversible - if U is unitary then so is U^{-1} by 8.4. However, the same is not true of classical gates. For example, the binary **AND** gate, which returns 1 only if both inputs are 1 and zero otherwise is not a reversible gate. In order to construct the quantum analog of classical gates, we aim to make reversible classical gates.

9.2. **A Naive Approach.** In this section we describe a correct but naive approach to creating quantum gates to perform any classical computation, albeit not efficiently. It is well known that the set $\{\mathbf{AND}, \mathbf{NOT}\}$ of Boolean gates is *functionally complete*, meaning that any Boolean computation can be done using just these two gates.

Let $\neg x$ denote the classical not function. Notice that this function is already reversible, and in fact we have already described its quantum parallel, the Pauli Transformation X.

Let $x \wedge y$ denote the Boolean **AND** operator. Consider what happens when we plug in the qubits $|x\rangle$, $|y\rangle$, $|0\rangle$ into the (reversible) Toffoli gate. We have

$$T|x, y, 0\rangle = |x, y, x \wedge y\rangle$$

In this way, by simply ignoring the first two kets of the result and only using the third, along with the X gate, we can compute anything a classical computer can.

9.3. **A Summary of a More Efficient Approach.** The major problem with the above naive approach is that it uses far too much space - every time something we perform the **AND** operation, we waste two qubits. The key is that we can undo operations after they have served their purpose, and then use them again. If we do this in the right way, then it possible to construct quantum circuit with similar efficiency as their classical counterpart.

Theorem 9.1. *Specifically, if any classical circuit uses t gates and s bits, it has a quantum (reversible) counterpart using only $O(t^{1+\epsilon})$ gates and $O(s \log t)$ qubits. We do not prove this result here, but a proof is given in [RP14].*

10. The Fourier Transform in Quantum Computation

10.1. **Discrete Fourier Transform.**

Definition 10.1. The *Discrete Fourier Transform* (DFT), maps $a(x)$ to $A(x)$, where $a : [0, 1, \ldots N-1] \to \mathbb{C}$ and $A : [0, 1, \ldots N-1] \to \mathbb{C}$ are discrete, complex valued functions. It is given by

$$A(x) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a(k) \exp\left(2\pi i \frac{kx}{N}\right)$$

Consider the Discrete Fourier Transform of a which has a frequency dividing N. Say

$$a(x) = \exp\left(-2\pi i \frac{ux}{N}\right), \tag{10.1}$$

where $0 \leq u < N$. Then, we have

$$\begin{aligned} A(x) &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a(k) \exp\left(2\pi i \frac{kx}{N}\right) \\ &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \exp\left(-2\pi i \frac{uk}{N}\right) \exp\left(2\pi i \frac{kx}{N}\right) \\ &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \exp\left(2\pi i \frac{k(u-x)}{N}\right) \end{aligned}$$

We can see that

$$\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \exp\left(2\pi i k \frac{r}{N}\right) = 0$$

if $r \not\equiv 0 \pmod{N}$ by the roots of unity. However, it is clear that if $r \equiv 0 \pmod{N}$, we have that each term in the sum is simply 1 and we get

$$\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \exp\left(2\pi i k \frac{r}{N}\right) = \frac{1}{\sqrt{N}} N = \sqrt{N}$$

So, for $0 \leq x < N$, the only x such that $A(x) \neq 0$ is $x = u$. That is,

$$A(x) = \begin{cases} 0 & \text{if } x \neq u \\ \sqrt{N} & \text{if } x = u. \end{cases}$$

Any complex-valued function with period r and frequency N/r can be approximated using its Fourier series as the sum of exponential functions whose frequencies are multiples of u. By the linearity of the DFT, we can think of the computation a linear combination of Fourier transforms of functions of the form of 10.1. If N divides r then as we saw, the only nonzero $A(x)$ will be multiples of $u = N/r$. According to [Sho97], if N does not divide r, we get only an approximation of this behavior - the highest values of $A(x)$ are when x is close to a multiple of $u = N/r$. The proof of this is somewhat involved, so we do not prove it here.

10.2. **The Quantum Fourier Transform.** The *quantum Fourier transform* (QFT) is very similar to the DFT. It is defined as

$$U_f : \sum_x a(x) |x\rangle \to \sum_x A(x) |x\rangle,$$

where $A(x)$ is defined from the DFT.

Lemma 10.2. *We claim that the matrix U_f defined above is unitary.*

10.3. **The Quantum Fast Fourier Transform.** It is possible to construct U_f in only $O(n^2)$ gates. We don't go into the details here, see [RP14].

11. Shor's Algorithm

Shor's algorithm is a quantum algorithm which factors numbers very efficiently. To be precise, Shor's algorithm runs in $O(n^2 \log n \log\log n)$ time, where n is the number of bits of the number that we want to factor. All current classical factoring algorithms are super-polynomial in n. This fact is the basis for most cryptographic algorithms today - this is discussed further in the conclusion. This section of the paper assumes a basic understanding of number theory.

11.1. **From Factoring to Period-Finding.** Consider the function $f(k) \equiv a^k \pmod M$. M is the number we want to factor. Let the period of f be r. Then, $f(r) = f(0)$, or $a^r \equiv 1 \pmod M$. If r is even, we can write

$$(a^{r/2} - 1)(a^{r/2} + 1) \equiv 0 \pmod M.$$

Then, notice that if $a^{r/2} + 1 \not\equiv 0 \pmod M$, computing $\gcd(a^{r/2} + 1, M)$ will give a nontrivial factor of M. Thus, we have converted the problem of factoring M into the problem of finding the period of f. And, we already have a tool which can help us find the period of a function - the Quantum Fourier Transform. The general outline of the algorithm so far looks like this:

(1) Randomly choose a and determine the period of $f(k) = a^k \pmod M$.
(2) if r is even, use Euclid's GCD algorithm to efficiently find $\gcd(a^{r/2} + 1, M)$.
(3) Repeat if nontrivial factor not found.

11.2. **The Quantum Part of the Algorithm.** The only part of the problem where we really need quantum computing is the first step. Consider the qubit $|00\ldots0\rangle\,|00\ldots0\rangle$, where each ket represents a state of n qubits. Then, apply the function $H \otimes H \otimes \cdots \otimes H$ (with n H's) to each of the kets. Recalling that $H\,|0\rangle = |+\rangle$ and the expansion of $|+\rangle\,|+\rangle \ldots |+\rangle$, we see that this state can also be written as

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle\,.$$

Then we can add another n qubits each in the $|0\rangle$ state:

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle\,|0\rangle\,.$$

where $N = 2^n$. Now we can apply $F : |x\rangle\,|0\rangle \to |x\rangle\,|f(x)\rangle$ to the second ket, so now we have

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle\,|f(x)\rangle$$

Now, one way to continue is to measure the second qubit. We will get some random value u for $f(x)$. Then, the state becomes

$$C \sum_{x=0}^{N-1} g(x)\,|x\rangle\,|u\rangle$$

where C is the appropriate scale factor and

$$g(x) = \begin{cases} 1 & \text{if } f(x) = u \\ 0 & \text{otherwise.} \end{cases}$$

Notice that the period of g is the same as the period of f. From here on out we can ignore the second ket $|u\rangle$ because it is no longer entangled with the first one. If we could just measure g twice and get consecutive values of x so that $g(x) = 1$. But, by the no-cloning theorem (8.5), we can't measure g twice. This is where we apply the quantum Fourier transform:

$$U_F(C \sum_{x=0}^{N-1} g(x) |x\rangle) = C' \sum_{c=0}^{N-1} G(c) |c\rangle ,$$

where $G(c) = \sum_x g(x) \exp(\frac{2\pi icx}{N})$. Recall that when the period r if g is a power of 2, then $G(c) \neq 0$ only when c is a multiple of N/r. Thus, when we measure, we will obtain the state $|v\rangle$ corresponding to a value v which is close to a multiple of N/r.

11.3. **Finding the factor of M.** Let's begin this section with a simple but useful lemma:

Lemma 11.1. *The absolute difference between two distinct fractions with denominators less than n is greater than $\frac{1}{n^2}$.*

Proof. Let the fractions be $\frac{a}{b}$ and $\frac{c}{d}$. Then we have

$$\left|\frac{a}{b} - \frac{c}{d}\right| = \left|\frac{ad - bc}{bd}\right| \geq \left|\frac{1}{bd}\right| > \frac{1}{n^2}$$

where the last inequality is strict because clearly the lemma holds if $b = d$. ■

Now we can obtain r with high probability using v. [Sho97] shows that with high probability

$$\left|v - j\frac{N}{r}\right| < \frac{1}{2}.$$

Recalling that $M^2 \leq N$, we have

$$\left|\frac{v}{N} - \frac{j}{r}\right| < \frac{1}{2N} \leq \frac{1}{2M^2}.$$

Then, we can evaluate $\frac{v}{N}$ and use the *continued fractions algorithm* to find the best fractional approximation $\frac{b}{c}$ for $\frac{v}{N}$ with $b, c < N$. By 11.1, any two fractions of the form $\frac{b}{c}$ form differ by more than $\frac{1}{N^2} > \frac{1}{2M^2}$, so the value of b/c that we find must be equal to j/r. Thus, have found r - it is just c.

Now that we have r, as described earlier, we can factor M by computing $\gcd(a^{r/2} + 1, M)$ and checking if this number is a nontrivial divisor of M.

11.4. **Complexity of Shor's Algorithm.** The complexity one iteration of Shor's Algorithm is $O(n^2 \log n \log\log n)$. No classical factoring algorithm even comes close to this - the current best classical factoring algorithm, the general number field sieve, runs in

$$O\left(\exp\left(((64/9)^{\frac{1}{3}} + o(1))(\ln n)^{\frac{1}{3}}(\ln\ln n)^{\frac{2}{3}}\right)\right),$$

which is superpolynomial. You might be wondering what exactly we mean when we say "with high probability". Shor showed that if we run Shor's algorithm only $\log\log r$ times, which in practice is a very small number, we are expected to get a factor of M.

12. Conclusion

12.1. **A Related Problem: The Discrete Logarithm Problem.** The discrete logarithm problem (DLP) is the problem of finding x given

$$a^x \equiv b \pmod{m}.$$

This problem, similarly to factoring, is also computationally difficult on a classical computer. There are modifications to Shor's algorithm which can solve the DLP much faster than any classical algorithm.

12.2. **A Generalization of Shor's Algorithm.** The following problem is a generalization of the problem of factoring, period finding, and the discrete logarithm:

Definition 12.1. *Finite Abelian Hidden Subgroup Problem:* Let G be a finite Abelian group with cyclic decomposition $G = \mathbb{Z}_{n_0} \times \cdots \times \mathbb{Z}_{n_L}$. Suppose G contains a subgroup H that is implicitly defined by a function f on G which is constant and distinct on every coset of H. Find a set of generators for H

Example. Let's see how we can think of the problem of period-finding as a hidden subgroup problem. Let f be a function on $\mathbb{Z}_N$ with period r. The subgroup $H \in \mathbb{Z}_N$ generated by r is the hidden subgroup. Then, once a generator for H is given, r can be found by computing $\gcd(r, N)$.

In addition to period finding, the discrete logarithm problem can also be viewed as a finite Abelian subgroup problem. The problem which Simon's algorithm (another quantum algorithm) solves can also be thought of as a finite Abelian subgroup problem. As you might have guessed, we have solved this problem on a quantum computer.

However, no one knows how to solve the corresponding problem for non-abelian groups, although some progress has been made.

12.3. **Cryptographic implications of Shor's Algorithm.** There are many cryptographic algorithms which rely on the difficulty of factoring or of the similar Discrete Logarithm Problem. A few of them are

(1) RSA
(2) Diffie-Hellman Key Exchange
(3) Digital Signature Algorithm

Quantum computers right now aren't extremely powerful - the largest number they can factor using Shor's Algorithm is only 21. However, quantum computers are improving rapidly, and even now, people are recording encrypted data with the goal of decrpyting them later using a quantum computer. For those reasons, there is an emerging field of *post-quantum cryptography*, which is the field of creating cryptographic algorithms which are resistant to quantum computers. If you'd like to learn more about Shor's algorithm or quantum computing in general, feel free to check out some of the resources in the bibliography.

References

[DN05] Christopher M. Dawson and Michael A. Nielsen. The solovay-kitaev algorithm, 2005.
[RP14] Eleanor Rieffel and Wolfgang Polak. *Quantum Computing: A gentle introduction*. The MIT Press, 2014.
[Sho97] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, oct 1997.

Email address: manu.isaacs@gmail.com