2002

Quantum factoring

Amy D. Vu
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Vu, Amy D., "Quantum factoring" (2002). *Master's Theses*. 2305.
DOI: <https://doi.org/10.31979/etd.2znj-vt4z>
https://scholarworks.sjsu.edu/etd_theses/2305

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

9

1

QUANTUM FACTORING

A Thesis

Presented to

The Faculty of the Department of Mathematics

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Amy D. Vu

May 2002

UMI Number: 1408822

UMI[®]

UMI Microform 1408822

**Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.**

**ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346**

© 2002

Amy D. Vu

ALL RIGHTS RESERVED

APPROVED FOR THE DEPARTMENT OF MATHEMATICS

Dr. Tim Hsu

Dr. Brian Peterson

Dr. Chris Pollett

APPROVED FOR THE UNIVERSITY

ABSTRACT

QUANTUM FACTORING

by Amy D. Vu

This thesis examines the topic of quantum factoring. In the first part, after review of several fundamental linear algebra concepts, the postulates of quantum mechanics and the Stern-Gerlach experiment are presented. The second part of this thesis focuses on quantum gates and the quantum Fourier transform. The key ideas from number theory that underlie quantum factoring are also presented in the second part. Finally, the quantum factoring algorithm is constructed using the quantum Fourier transform and the results in number theory. In particular, we explain why quantum factoring is exponentially faster than any known classical factoring algorithm.

TABLE OF CONTENTS

1	Introduction	1
2	Linear Algebra	3
3	Postulates of Quantum Mechanics	17
4	Stern-Gerlach Experiment	26
5	Quantum Gates and Circuits	31
6	Quantum Fourier Transform	39
7	Number Theory	50
8	Quantum Factoring Algorithm	63
	References	73

1 Introduction

In 1982, the Nobel-prize physicist Richard Feynman suggested that simulation of a quantum mechanical system on a classical computer is intrinsically complex and costly and these difficulties could be avoided by building a quantum machine based on quantum mechanics rather than classical mechanics. For some time, the notion of a quantum computer was primarily of theoretical interest. More recent developments like the factoring algorithm have scientists racing to build the first useable quantum computer.

The most fundamental building block in a classical computer is a bit. The bit can only exist in two distinct states 0 or 1. The building block of a quantum computer is called a quantum bit, or qubit. Roughly speaking, the qubit can exist in the classical 0, 1 state as well as in a superposition of the two states. When a qubit is in a superposition of the states, it is thought to exist in both the 0 state and the 1 state. Thus, an operation on a qubit in a superposition of states is an operation on both the 0 bit and the 1 bit at the same time. This implies that by applying an operation on a single qubit, we have performed a simultaneous operation on two different values. In a similar fashion, a two-qubit system would perform the operation on 4 values; a three-qubit system would perform the operation on 8 values; and an n -qubit system would perform the operation on 2^n values. Increasing the number of qubits in a system exponentially increases the quantum parallelism, the number of values that can be evaluated at the same time, which can be obtained with the system.

Once a calculation using quantum parallelism has been performed, the results are given in many different states. Again, roughly speaking, a solution to a computation can

only be obtained by examining the interference of the various results. However, once a result has been examined, or measured, no further quantum computations may take place.

As far as we know, although the quantum computer can exploit the theory of quantum parallelism, the quantum computer cannot outperform the classical computer in all computational tasks. Nevertheless, the factoring algorithm conjured up by Peter Shor in 1994, has been shown to be exponentially faster than the best classical factoring methods. So, a quantum computer using Shor's algorithm would be able to factor numbers with large prime factors very quickly. Hence, it would be able to break some of the current cryptology methods in a matter of seconds.

2 Linear Algebra

We begin our study of quantum mechanics with some theory from linear algebra. Unless otherwise cited, most of the theorems presented in this section are found Horn and Johnson [HoJo].

Definition 2.1: The *adjoint* † of a matrix A is the transpose of the complex conjugate * of A , i.e. $A^\dagger = (A^*)^\top$. Similarly, $\bar{v}^\dagger = (\bar{v}^*)^\top$ for a vector $\bar{v}$.

Definition 2.2: Let V be a vector space over $\mathbb{C}$. An *inner product* on V is a function $\langle \cdot, \cdot \rangle$ from $V \times V$ to $\mathbb{C}$ such that for any $\bar{x}, \bar{y}, \bar{z} \in V$ and $c \in \mathbb{C}$, we have

- (i) $\langle \bar{x} + \bar{z}, \bar{y} \rangle = \langle \bar{x}, \bar{y} \rangle + \langle \bar{z}, \bar{y} \rangle$.
- (ii) $\langle c\bar{x}, \bar{y} \rangle = c^* \langle \bar{x}, \bar{y} \rangle$ where * denotes complex conjugation.
- (iii) $\langle \bar{x}, \bar{y} \rangle^* = \langle \bar{y}, \bar{x} \rangle$.
- (iv) $\langle \bar{x}, \bar{x} \rangle > 0$ if $\bar{x} \neq 0$.

Example 2.3: Define a function $\langle \cdot, \cdot \rangle$ from $\mathbb{C}^n \rightarrow \mathbb{C}$ by $\langle \bar{v}, \bar{w} \rangle = \bar{v}^\dagger \bar{w}$. $\langle \cdot, \cdot \rangle$ is an inner product since

- (i) $\langle \bar{x} + \bar{z}, \bar{y} \rangle = (\bar{x} + \bar{z})^\dagger \bar{y} = (\bar{x}^\dagger + \bar{z}^\dagger) \bar{y} = \bar{x}^\dagger \bar{y} + \bar{z}^\dagger \bar{y} = \langle \bar{x}, \bar{y} \rangle + \langle \bar{z}, \bar{y} \rangle$.
- (ii) $\langle c\bar{x}, \bar{y} \rangle = (c\bar{x})^\dagger \bar{y} = (c^* \bar{x}^\dagger) \bar{y} = c^* (\bar{x}^\dagger \bar{y}) = c^* \langle \bar{x}, \bar{y} \rangle$.
- (iii) $\langle \bar{x}, \bar{y} \rangle^* = (\bar{x}^\dagger \bar{y})^* = \bar{y}^\dagger \bar{x} = \langle \bar{y}, \bar{x} \rangle$.

$$(iv) \quad \langle \bar{x}, \bar{x} \rangle^\dagger = \bar{x}^\dagger \bar{x} = \sum_i x_i^* x_i = \sum_i |x_i|^2 \geq 0 \text{ for all } x \text{ and } \langle \bar{x}, \bar{x} \rangle = 0 \text{ if and only if}$$

$$\text{each } x_i = 0.$$

A vector space equipped with an inner product $\langle \cdot, \cdot \rangle$ is called an *inner product space*. If H is an inner product space such that the norm defined by

$$\|\bar{v}\| = \sqrt{\langle \bar{v}, \bar{v} \rangle}$$

turns H into a complete metric space (a metric space in which every Cauchy sequence is convergent), then we say H is a *Hilbert space*. For finite-dimensional vector spaces over $\mathbb{C}$, an inner product space is exactly the same as a Hilbert space.

The inner product described in Example 2.3 is the standard inner product on $\mathbb{C}^n$. For the rest of this paper, $\langle \cdot, \cdot \rangle$ will be used to denote this standard inner product.

Definition 2.4: A matrix $A \in M_n(\mathbb{C})$ is said to be *unitary* if $A^\dagger A = I$. The set of all unitary matrices $U_n(\mathbb{C})$ is a subgroup of $GL_n(\mathbb{C})$, the group of invertible $n \times n$ matrices. If $A = A^\dagger$, then we say A is *Hermitian*.

A set of vectors $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n$ is said to be *orthonormal* if each v_i in the set is a unit vector, i.e., $\|v_i\| = 1$, and for any v_i, v_j in the set, $i \neq j$, we have $\langle v_i, v_j \rangle = 0$. If $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n$ is linearly independent, but not necessarily orthonormal, we can apply the Gram-Schmidt orthonormalization process to form an orthonormal set of vectors with the same span. Moreover, if $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n$ forms a basis for $\mathbb{C}^n$, then the vectors

constructed using this process forms an orthonormal basis for $\mathbb{C}^n$. Recall that the Gram-Schmidt orthonormalization process can be also used to extend any orthonormal set into an orthonormal basis [HoJo].

Theorem 2.5 [FIS]: A linear operator U on a finite-dimensional space V is unitary if and only if for some orthonormal basis β of V , $U(\beta)$ is an orthonormal basis for V .

Definition 2.6: A matrix $B \in M_n(\mathbb{C})$ is *unitarily equivalent* to $A \in M_n(\mathbb{C})$ if there is a unitary matrix $U \in M_n(\mathbb{C})$ such that $B = U^\dagger A U$.

One very important theorem regarding unitary equivalence is Schur's Triangularization Theorem given below. We use the notation $\sigma(A)$ to denote the spectrum of A , the set of eigenvalues of a matrix A .

Theorem 2.7 (Schur's Theorem): Given $A \in M_n(\mathbb{C})$, there exists a unitary matrix $U \in M_n(\mathbb{C})$ such that $U^\dagger A U = T$, where T is an upper triangular matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ in an arbitrary predetermined order.

Proof: We will prove this by induction on n . The case when $n = 1$ is trivial. Assume that any $k \times k$ complex matrix is unitarily equivalent to a triangular matrix with a predetermined order of its diagonal elements. Consider $A \in M_{k+1}(\mathbb{C})$ with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_{k+1}$ in a prescribed order. Let $\bar{u}_1$ be a unit eigenvector of A with

eigenvalue λ_1 . We can extend it to an orthonormal basis $\{\bar{u}_1, \bar{u}_2, \dots, \bar{u}_{k+1}\}$ of $\mathbb{C}^{k+1}$. Let

$S = [\bar{u}_1 \ \bar{u}_2 \ \dots \ \bar{u}_{k+1}]$. Then

$$\begin{aligned} AS &= [A\bar{u}_1 \ A\bar{u}_2 \ \dots \ A\bar{u}_{k+1}] \\ &= [\lambda_1 \bar{u}_1 \ A\bar{u}_2 \ \dots \ A\bar{u}_{k+1}]. \end{aligned}$$

Since $S^\dagger = S^{-1}$ implies that $S^\dagger(\lambda \bar{u}_i) = \lambda(S^\dagger \bar{u}_i) = \lambda \bar{e}_i$, we have

$$S^\dagger AS = \begin{bmatrix} \lambda_1 & * \\ 0 & A_2 \end{bmatrix}$$

where A_2 is a $k \times k$ matrix with eigenvalues $\lambda_2, \lambda_2, \dots, \lambda_{k+1}$. A_2 has eigenvalues

$\lambda_2, \lambda_3, \dots, \lambda_{k+1}$ because

$$\sigma(A) = \sigma(S^\dagger AS) = \sigma\left(\begin{bmatrix} \lambda_1 & * \\ 0 & A_2 \end{bmatrix}\right) = \{\lambda_1\} \cup \sigma(A_2).$$

By our induction assumption, there exists a unitary matrix S_2 such that

$$S_2^\dagger A_2 S_2 = \begin{bmatrix} \lambda_2 & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_{k+1} \end{bmatrix}.$$

Hence, if $S_1 = \begin{bmatrix} 1 & 0 \\ 0 & S_2 \end{bmatrix}$,

$$\begin{aligned} S_1^\dagger S^\dagger A S S_1 &= \begin{bmatrix} 1 & 0 \\ 0 & S_2^\dagger \end{bmatrix} \begin{bmatrix} \lambda_1 & * \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & S_2 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & * \\ 0 & S_2^\dagger (A_2) S_2 \end{bmatrix} \\ &= T, \end{aligned}$$

where T is an upper triangular matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ in an arbitrary

predetermined order for all n . Note that $\begin{bmatrix} 1 & 0 \\ 0 & S_2 \end{bmatrix} = S_1$ is unitary since

$$\begin{bmatrix} 1 & 0 \\ 0 & S_2 \end{bmatrix}^\dagger = \begin{bmatrix} 1 & 0 \\ 0 & S_2^\dagger \end{bmatrix}, \text{ so } \begin{bmatrix} 1 & 0 \\ 0 & S_2 \end{bmatrix}^\dagger \begin{bmatrix} 1 & 0 \\ 0 & S_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I_k \end{bmatrix}. \text{ Therefore, } \tilde{S} = SS_1 \text{ is unitary}$$

because $U_n(\mathbb{C})$ is a subgroup of $GL_n(\mathbb{C})$. $\square$

Definition 2.8: If $A \in M_n(\mathbb{C})$ is unitarily equivalent to a diagonal matrix, then A is *unitarily diagonalizable*.

Definition 2.9: A matrix $A \in M_n(\mathbb{C})$ is *normal* if $A^\dagger A = AA^\dagger$.

Proposition 2.10: If U is either unitary or Hermitian (or skew-Hermitian), then U is normal.

Proof: If U is unitary, then $U^\dagger U = I = UU^\dagger$, so U is normal. On the other hand, if U is Hermitian, then we have $U^\dagger U = UU^\dagger$ because $U = U^\dagger$. $\square$

Theorem 2.11: If $H \in M_n(\mathbb{C})$ is a Hermitian matrix, then all of the eigenvalues of H are real.

Proof: If H is a Hermitian matrix, then $(\bar{x}^\dagger H \bar{x})^* = (\bar{x}^\dagger H \bar{x})^\dagger = \bar{x}^\dagger H^\dagger \bar{x} = \bar{x}^\dagger H \bar{x}$ for all

$\bar{x} \in \mathbb{C}^n$. This implies that $\bar{x}^\dagger H \bar{x}$ is real for all $\bar{x} \in \mathbb{C}^n$ since it equals its complex conjugate. If λ is any arbitrary eigenvalue of H , then there exists a nonzero eigenvector

$\bar{x}$ with respect to λ such that $H\bar{x} = \lambda\bar{x}$ and $\bar{x}^\dagger \bar{x} = 1$. Thus, $\lambda = \lambda\bar{x}^\dagger \bar{x} = \bar{x}^\dagger \lambda\bar{x} = \bar{x}^\dagger H\bar{x}$ is real. $\square$

One important class of Hermitian operators is the projectors. For any n -dimensional vector space V and k -dimensional subspace W of V , we can extend an orthonormal basis $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_k$ of W to an orthonormal basis

$\bar{v}_1, \bar{v}_2, \dots, \bar{v}_k, \bar{v}_{k+1}, \dots, \bar{v}_n$ of V . We say $P = \sum_{i=1}^k \bar{v}_i \bar{v}_i^\dagger$ is the *projector* onto W . Notice that $\bar{v}_i \bar{v}_i^\dagger$ is Hermitian for any $1 \leq i \leq n$, so P is Hermitian. The orthogonal complement of P is defined to be $Q = I - P$.

Lemma 2.12: An upper triangular normal matrix is a diagonal matrix.

Proof: Let T be an upper triangular normal matrix. Since $T^\dagger T = TT^\dagger$, the (i, i) entry of $T^\dagger T$ is the same as the (i, i) entry of $TT^\dagger$, and so we have

$$t_{ii}^* t_{ii} = t_{ii} t_{ii}^* + \sum_{j=i+1}^n t_{ij} t_{ij}^* = |t_{ii}|^2 + \sum_{j=i+1}^n |t_{ij}|^2,$$

implying that $0 = \sum_{j=i+1}^n |t_{ij}|^2$. This states that the sum of a set of nonnegative terms is equal

to zero. It follows that for $j \geq i+1$, t_{ij} is equal to zero. Therefore, T is a diagonal matrix. $\square$

Lemma 2.13: If A is a normal matrix and T is a matrix such that $A = U^\dagger T U$, where U is a unitary matrix, then T is normal.

Proof: From the definition of normal and unitary, we get

$$\begin{aligned}
TT^\dagger &= (U^\dagger AU)(U^\dagger AU)^\dagger = U^\dagger AUU^\dagger A^\dagger U = U^\dagger AA^\dagger U = U^\dagger A^\dagger AU \\
&= U^\dagger A^\dagger UU^\dagger AU = (U^\dagger AU)^\dagger (U^\dagger AU) = T^\dagger T. \quad \square
\end{aligned}$$

Theorem 2.14: A matrix $A = [a_{ij}] \in M_n(\mathbb{C})$ is normal if and only if A is unitarily diagonalizable.

Proof: Suppose A is normal. By Schur's Theorem (2.7), there exists an upper triangular matrix $T = [t_{ij}] \in M_n(\mathbb{C})$ that is unitarily equivalent to A . Since A is normal, T must be normal (2.13), but any upper triangular normal matrix must be diagonal (2.12).

Therefore, A is unitarily diagonalizable. Conversely, since diagonal matrices are normal and unitary equivalence preserves normality, A is normal. $\square$

Theorem 2.15: If A and B are two commuting normal matrices, then A and B are simultaneously unitarily diagonalizable.

Proof: By Theorem 2.14, we can find a unitary matrix U such that

$$U^\dagger AU = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix},$$

where $\lambda_1, \lambda_2, \dots, \lambda_n$ are eigenvalues of A . Let $B_1 = U^\dagger BU = [b_{ij}]$. Since A and B

commute, Λ and B_1 must also commute; hence $\lambda_i b_{ij} = b_{ij} \lambda_j$. So, $(\lambda_i - \lambda_j) b_{ij} = 0$,

implying that $b_{ij} = 0$ whenever $\lambda_i \neq \lambda_j$. We now change notation so that $\lambda_1, \lambda_2, \dots, \lambda_k$ are distinct eigenvalues of A and assume they are ordered so that

$$U^{\dagger}AU = \begin{bmatrix} \lambda_1 I & & 0 \\ & \ddots & \\ 0 & & \lambda_k I \end{bmatrix},$$

where each $\lambda_i I$ is a scalar matrix whose size equals the multiplicity of λ_i . It follows that

$$B = \begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_k \end{bmatrix}$$

is a block diagonal matrix where each B_i block has the same size as $\lambda_i I$. Since B is

unitarily diagonalizable, B is normal, so each B_i is normal and thus, B_i is unitarily

diagonalizable (2.13). Therefore, we can find a unitary T_i such that $T_i^{\dagger} B_i T_i$ is diagonal.

Now, we have

$$U^{\dagger}AU = \begin{bmatrix} \lambda_1 I & & 0 \\ & \ddots & \\ 0 & & \lambda_k I \end{bmatrix}.$$

For

$$T = \begin{bmatrix} T_1 & & 0 \\ & \ddots & \\ 0 & & T_k \end{bmatrix},$$

we get that T is also unitary, and $T^{\dagger} B T$ is diagonal. Also, $T_i^{\dagger} \lambda_i I T_i = \lambda_i (T_i^{\dagger} T) = \lambda_i I$ so

$T^{\dagger} A T$ is diagonal. Therefore, A and B are simultaneously unitarily

diagonalizable. $\square$

For any matrix $A \in M_n(\mathbb{C})$, define $e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots$. Recall that the solution to the differential equation $\frac{dx}{dt} = Ax$ is $x = e^{At}x_0$. From analysis, we know that the sequence of partial sums of e^A is convergent. Notice that for a unitary matrix U and positive integer n , $(U^\dagger AU)^n = U^\dagger A^n U$. Hence,

$$\begin{aligned} e^{U^\dagger AU} &= I + U^\dagger AU + \frac{(U^\dagger AU)^2}{2!} + \frac{(U^\dagger AU)^3}{3!} + \dots \\ &= I + U^\dagger AU + \frac{U^\dagger A^2 U}{2!} + \frac{U^\dagger A^3 U}{3!} + \dots \\ &= U^\dagger \left(I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots \right) U \\ &= U^\dagger e^A U. \end{aligned}$$

Also, note that if $D = \text{diag}(a_1, a_2, \dots, a_n)$ is a diagonal matrix, then

$$\begin{aligned} e^D &= I + D + \frac{D^2}{2!} + \frac{D^3}{3!} + \dots \\ &= \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} + \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix} + \frac{1}{2!} \begin{bmatrix} a_1^2 & & \\ & \ddots & \\ & & a_n^2 \end{bmatrix} + \dots \\ e^D &= \begin{bmatrix} e^{a_1} & & \\ & \ddots & \\ & & e^{a_n} \end{bmatrix}. \end{aligned}$$

Corollary 2.16: Suppose A and B are commuting normal matrices. Then $e^A e^B = e^{A+B}$.

Proof: Since A and B are normal, it is easy to see that e^A and e^B are also normal and e^{A+B} commutes with e^A and e^B . By Theorem 2.14, e^A and e^B are commuting matrices,

so e^A and e^B are simultaneously unitarily diagonalizable, i.e., there exists a unitary U such that $U^\dagger e^A U = \Lambda_1$ and $U^\dagger e^B U = \Lambda_2$ where Λ_1 and Λ_2 are diagonal matrices.

Therefore, $e^A e^B = (U \Lambda_1 U^\dagger)(U \Lambda_2 U^\dagger) = U \Lambda_1 \Lambda_2 U^\dagger = e^{A+B}$. $\square$

We now turn our attention to the outer product of a vector space. Given two inner product spaces V and W , an *outer product* on V is a linear transformation $\xi_{v,w} : V \rightarrow W$ whose action on an arbitrary $\bar{x} \in V$ is defined by

$$\xi_{v,w}(\bar{x}) = \bar{w} \langle \bar{v}, \bar{x} \rangle = \langle \bar{v}, \bar{x} \rangle \bar{w}$$

for fixed $\bar{v} \in V$ and $\bar{w} \in W$. Notice that the matrix of $\xi_{v,w}$ is $\overline{wv}^\dagger$.

One very important result stemming from the notion of an outer product is the *completeness relation for orthonormal vectors*. Suppose that $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n$ is an orthonormal basis for the inner product space V . By definition of basis, this implies that any vector $v \in V$ can be written as $\sum_i a_i v_i$ where each $a_i \in \mathbb{C}$. It should be noted that $\sum_j a_j \langle \bar{v}_i, \bar{v}_j \rangle = a_i$ since $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n$ is an orthonormal set of vectors. Therefore, for $\bar{v} = \sum_j a_j \bar{v}_j$,

$$\begin{aligned} (\sum_i \xi_{v_i, v_i})(\bar{v}) &= (\sum_i \bar{v}_i \langle \bar{v}_i, \bar{v} \rangle) \\ &= \sum_i \bar{v}_i \langle \bar{v}_i, \sum_j a_j \bar{v}_j \rangle \\ &= \sum_i \bar{v}_i \sum_j a_j \langle \bar{v}_i, \bar{v}_j \rangle \\ &= \sum_i a_i \bar{v}_i \\ &= \bar{v}. \end{aligned}$$

It follows that

$$\sum_i (\bar{v}_i)(\bar{v}_i)^\dagger = I. \quad (2.16)$$

Equation 2.16 is called the *completeness relation*.

Another important concept that we will need is the tensor product of vector spaces. For the purpose of this paper, we will present an informal definition of a tensor product of vectors and vector spaces by giving a brief description these elements. (The formal definition of tensor products can be found in Hungerford [Hung].)

Let V be an n -dimensional vector space and W be an m -dimensional vector space. Then $V \otimes W$ is a vector space, and for any $\bar{v} \in V$ and $\bar{w} \in W$, $\bar{v} \otimes \bar{w}$ is a vector in $V \otimes W$ such that for any $r \in \mathbb{C}$, $\bar{v}$, $\bar{v}_1$, $\bar{v}_2 \in V$ and $\bar{w}$, $\bar{w}_1$, $\bar{w}_2 \in W$, we have the following properties:

- (i) $(\bar{v}_1 + \bar{v}_2) \otimes \bar{w} = (\bar{v}_1 \otimes \bar{w}) + (\bar{v}_2 \otimes \bar{w}).$
- (ii) $\bar{v} \otimes (\bar{w}_1 + \bar{w}_2) = (\bar{v} \otimes \bar{w}_1) + (\bar{v} \otimes \bar{w}_2).$
- (iii) $r(\bar{v} \otimes \bar{w}) = r\bar{v} \otimes \bar{w} = \bar{v} \otimes r\bar{w}.$

The next proposition regarding the basis of $V \otimes W$ will be given without proof, since the proof requires a more formal definition of tensor products.

Proposition 2.17: If V is an n -dimensional vector space with basis $\beta_1 = \{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n\}$

and W is an m -dimensional vector space with basis $\beta_2 = \{\bar{w}_1, \bar{w}_2, \dots, \bar{w}_m\}$, then their

tensor product $V \otimes W$ is an nm -dimensional vector space with basis

$$\beta = \{\bar{v}_i \otimes \bar{w}_j: \bar{v}_i \in \beta_1 \text{ and } \bar{w}_j \in \beta_2\}.$$

This implies that if $\bar{v}$ has coordinate vector $[a_1 \ a_2 \ \dots \ a_n]^T$ relative to the basis $\{\bar{v}_i\}$ of V , and $\bar{w}$ has coordinate vector $[b_1 \ b_2 \ \dots \ b_m]^T$ relative to the basis $\{\bar{w}_j\}$ of W , then $\bar{v} \otimes \bar{w}$ has coordinate vector $[a_1 b_1 \ \dots \ a_n b_m]^T$ relative to the basis $\{\bar{v}_i \otimes \bar{w}_j\}$ of $V \otimes W$.

Proposition 2.18: If V and W are Hilbert spaces, then the function

$$\left\langle \sum a_{ij} (\bar{v}_i \otimes \bar{w}_j), \sum b_{kl} (\bar{v}_k \otimes \bar{w}_l) \right\rangle = \sum a_{ij}^* b_{kl} \langle \bar{v}_i, \bar{v}_k \rangle \langle \bar{w}_j, \bar{w}_l \rangle$$

is an inner product on $V \otimes W$.

Proof: We only need to verify that the function defined above satisfies the necessary conditions to be an inner product:

$$\begin{aligned} \text{(i)} \quad \langle \bar{x} + \bar{z}, \bar{y} \rangle &= \left\langle \sum a_{ij} (\bar{v}_i \otimes \bar{w}_j) + \sum c_{pq} (\bar{v}_p \otimes \bar{w}_q), \sum b_{kl} (\bar{v}_k \otimes \bar{w}_l) \right\rangle \\ &= \left\langle \sum (a_{ij} + c_{ij}) (\bar{v}_i \otimes \bar{w}_j), \sum b_{kl} (\bar{v}_k \otimes \bar{w}_l) \right\rangle \\ &= \sum a_{ij}^* b_{kl} \langle \bar{v}_i, \bar{v}_k \rangle \langle \bar{w}_j, \bar{w}_l \rangle + \sum c_{ij}^* b_{kl} \langle \bar{v}_i, \bar{v}_k \rangle \langle \bar{w}_j, \bar{w}_l \rangle \\ &= \langle \bar{x}, \bar{y} \rangle + \langle \bar{z}, \bar{y} \rangle \end{aligned}$$

$$\begin{aligned}
\text{(ii)} \quad \langle c\bar{x}, \bar{y} \rangle &= \left\langle c \sum a_{ij} (\bar{v}_i \otimes \bar{w}_j), \sum b_{kl} (\bar{v}_k \otimes \bar{w}_l) \right\rangle \\
&= \left\langle \sum a_{ij} (c\bar{v}_i \otimes \bar{w}_j), \sum b_{kl} (\bar{v}_k \otimes \bar{w}_l) \right\rangle \\
&= \sum a_{ij}^* b_{kl} \langle c\bar{v}_i, \bar{v}_k \rangle \langle \bar{w}_j, \bar{w}_l \rangle \\
&= c^* \sum a_{ij}^* b_{kl} \langle \bar{v}_i, \bar{v}_k \rangle \langle \bar{w}_j, \bar{w}_l \rangle \\
&= c^* \langle \bar{x}, \bar{y} \rangle
\end{aligned}$$

$$\begin{aligned}
\text{(iii)} \quad \langle \bar{x}, \bar{y} \rangle^* &= \left\langle \sum a_{ij} (\bar{v}_i \otimes \bar{w}_j), \sum b_{kl} (\bar{v}_k \otimes \bar{w}_l) \right\rangle^* \\
&= \left(\sum a_{ij}^* b_{kl} \langle \bar{v}_i, \bar{v}_k \rangle \langle \bar{w}_j, \bar{w}_l \rangle \right)^* \\
&= \sum (b_{kl}^* a_{ij}) \langle \bar{v}_k, \bar{v}_i \rangle \langle \bar{w}_l, \bar{w}_j \rangle \\
&= \langle \bar{y}, \bar{x} \rangle
\end{aligned}$$

$$\begin{aligned}
\text{(iv)} \quad \langle \bar{x}, \bar{x} \rangle &= \left\langle \sum a_{ij} (\bar{v}_i \otimes \bar{w}_j), \sum a_{pq} (\bar{v}_p \otimes \bar{w}_q) \right\rangle \\
&= \sum a_{ij}^* a_{pq} \langle \bar{v}_i, \bar{v}_p \rangle \langle \bar{w}_j, \bar{w}_q \rangle \\
&= \sum a_{ij}^* a_{ij} \\
&= \sum |a_{ij}|^2 \geq 0 \text{ if } \bar{x} \neq 0. \quad \square
\end{aligned}$$

It should be noted that if V and W are both inner product spaces over $\mathbb{C}$, then using the inner product defined in Proposition 2.18 on the space $V \otimes W$, we have that $V \otimes W$ is a finite dimensional inner product space. Thus, $V \otimes W$ is a Hilbert space.

Proposition 2.19: If $\{v_i\}$, $\{w_j\}$ are orthonormal basis for V and W respectively, then

$\{v_i \otimes w_j\}$ is an orthonormal basis for $V \otimes W$.

Proof: By Proposition 2.17, $\{v_i \otimes w_j\}$ is a basis for $V \otimes W$. For distinct elements

$v_k \otimes w_l$ and $v_p \otimes w_q$ in the basis of $V \otimes W$, we have

$$\langle v_k \otimes w_l, v_p \otimes w_q \rangle = \langle \bar{v}_k, \bar{v}_p \rangle \langle \bar{w}_l, \bar{w}_q \rangle = 0$$

since either $\bar{v}_k \neq \bar{v}_p$ or $\bar{w}_l \neq \bar{w}_q$, and therefore either $\langle \bar{v}_k, \bar{v}_p \rangle$ or $\langle \bar{w}_l, \bar{w}_q \rangle$ is 0. On the

other hand,

$$\langle v_k \otimes w_l, v_k \otimes w_l \rangle = \langle \bar{v}_k, \bar{v}_k \rangle \langle \bar{w}_l, \bar{w}_l \rangle = 1 \cdot 1 = 1$$

because $\{v_i\}$, $\{w_j\}$ are orthonormal bases for V and W . $\square$

3 Postulates of Quantum Mechanics

Quantum mechanics is the framework of quantum computation. It is the link from physical systems to mathematics. In this section, the foundations of quantum mechanics will be developed by exploring the four postulates of quantum mechanics, following the approach of Nielsen and Chuang [NiCh].

We begin this section by introducing some commonly used quantum mechanical notation:

$|v\rangle$ Vector in $\mathbb{C}^n$. Also known as a *ket*.

$\langle v|$ Adjoint of $|v\rangle$. Also known as a *bra*.

$\langle v|w\rangle$ Inner product of $|v\rangle$ and $|w\rangle$, i.e. $\langle v|w\rangle = |v\rangle^\dagger |w\rangle$.

Now, the first postulate of quantum mechanics establishes the state space, the vector space in which the quantum system takes place.

Postulate One: Every isolated physical system can be described by a finite-dimensional Hilbert space. The state vector of a quantum system is a unit vector in the state space. The vectors $|v\rangle$ and $\lambda|v\rangle$ represent the same physical state ($\lambda \in \mathbb{C}$ and $\lambda \neq 0$).

The system we are mainly concerned with is the *qubit* (quantum bit) system, the quantum system whose state space is $\mathbb{C}^2$. The vectors

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

form an orthonormal basis for $\mathbb{C}^2$. An arbitrary qubit of $\mathbb{C}^2$ therefore has the form

$$|v\rangle = a|0\rangle + b|1\rangle$$

with $|a|^2 + |b|^2 = 1$.

The second postulate of quantum mechanics gives us a description of the state vectors with respect to time. This postulate is presented in two forms.

Postulate Two (A): Let $|v\rangle$ be a state vector in a closed quantum system. The time evolution of this state vector is given by the Schrödinger equation

$$i\hbar \frac{d|v\rangle}{dt} = H|v\rangle$$

where $\hbar$ is a real number called Planck's constant and H is a fixed Hermitian matrix that governs the system.

Both the constant $\hbar$ and matrix H must be determined through experimentation. For the purpose of this paper, we will assume that $\hbar$ is known; H depends on the physical system in question.

Postulate Two (B): Given a state vector $|v\rangle$, the evolution of this state vector can be described by a unitary matrix on the state space, i.e.,

$$|v(t_2)\rangle = U|v(t_1)\rangle$$

where U is a unitary linear transformation and $|v(t_1)\rangle$, $|v(t_2)\rangle$ denote the state vector at time t_1 and t_2 respectively.

The connection between Postulate Two (A) and Postulate Two (B) lies in the solution to the Schrödinger equation (see page 11):

$$i\hbar \frac{d|v\rangle}{dt} = H|v\rangle$$

$$\frac{d|v\rangle}{dt} = \frac{1}{i\hbar} H|v\rangle$$

$$\frac{d|v\rangle}{dt} = \frac{-iH}{\hbar} |v\rangle$$

$$|v(t_2)\rangle = e^{\frac{-iH(t_2-t_1)}{\hbar}} |v(t_1)\rangle.$$

Proposition 3.1: If H is Hermitian, then $U = e^{\frac{-iH}{\hbar}}$ is a unitary matrix.

Proof: From the definition of Hermitian and Corollary 2.15, we have

$$\begin{aligned} U^\dagger U &= \left(e^{\frac{-iH}{\hbar}} \right)^\dagger \left(e^{\frac{-iH}{\hbar}} \right) = \left(e^{\frac{iH^\dagger}{\hbar}} \right) \left(e^{\frac{-iH}{\hbar}} \right) \\ &= e^{\frac{iH^\dagger}{\hbar} + \frac{-iH}{\hbar}} = e^{\frac{iH}{\hbar} + \frac{-iH}{\hbar}} = e^0 = I. \end{aligned}$$

Therefore, $U = e^{\frac{-iH}{\hbar}}$ is a unitary matrix. $\square$

Example 3.2: Using the Hermitian operator $H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, we have

$$\frac{-i}{\hbar} H = \frac{-i}{\hbar} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

with eigenvalues $-\frac{i}{\hbar}$ and $\frac{i}{\hbar}$. Note that $e^{\frac{-iH}{\hbar}t} = \begin{bmatrix} e^{(-\frac{i}{\hbar})t} & 0 \\ 0 & e^{(\frac{i}{\hbar})t} \end{bmatrix}$,

and so,

$$\left(\overline{e^{-\frac{i}{\hbar}Ht}}\right)^T \left(e^{-\frac{i}{\hbar}Ht}\right) = \begin{bmatrix} e^{(i/\hbar)t} & 0 \\ 0 & e^{(i/\hbar)t} \end{bmatrix} \begin{bmatrix} e^{(i/\hbar)t} & 0 \\ 0 & e^{(i/\hbar)t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

i.e., $\left(e^{-\frac{i}{\hbar}Ht}\right)$ is unitary. For the initial state vector $|v_0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, the time evolution

of this state vector is given by $|v_t\rangle = e^{-\frac{i}{\hbar}Ht} \left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\right)$. If $t = 2$, we get

$$|v_2\rangle = e^{-\frac{i}{\hbar}Ht} \left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\right) = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{-\frac{i}{\hbar}Ht} & 0 \\ 0 & e^{-\frac{i}{\hbar}Ht} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{-\frac{i}{\hbar}Ht} \\ e^{-\frac{i}{\hbar}Ht} \end{bmatrix}.$$

In terms of single qubits, it is possible for any unitary operator to be a description of an evolution of a qubit system, i.e., any 2×2 unitary operator can be physically realized.

The next postulate explains what happens when we interact with quantum systems. Since interaction with a quantum system means that we no longer have a closed quantum system, we must find a way to model the effects of measurements on quantum systems.

Definition 3.3: Let $\{M_m\}$ be a collection of matrices satisfying the equation

$\sum_n M_m^\dagger M_m = I$. Then $\{M_m\}$ is called quantum measurement and each M_m is called a

measurement operator. The index m represents the measurement outcomes that may occur in the experiment.

Theorem 3.4: Let $|0\rangle, |1\rangle, \dots, |n-1\rangle$ be an orthonormal basis for $\mathbb{C}^n$ and $M_i = |i\rangle\langle i|$.

Then $\{M_i\}$ is a quantum measurement. We say $M_i = |i\rangle\langle i|$ is a measurement with respect to the given basis $\{|i\rangle\}$.

Proof: It suffices to show that the collection $\{M_i\}$ satisfies the equation in Definition

3.3. Since $|0\rangle, |1\rangle, \dots, |n-1\rangle$ is a set of orthonormal vectors, it follows from the completeness equation (2.16) that $\sum_i |i\rangle\langle i| = I$. So,

$$\begin{aligned} \sum_i M_i^\dagger M_i &= \sum_i (|i\rangle\langle i|)^\dagger (|i\rangle\langle i|) \\ &= \sum_i (|i\rangle\langle i|)^\dagger (|i\rangle\langle i|) \\ &= \sum_i (|i\rangle\langle i|) (|i\rangle\langle i|) \\ &= \sum_i |i\rangle\langle i| |i\rangle\langle i| \\ &= \sum_i |i\rangle 1 \langle i| = I. \quad \square \end{aligned}$$

Postulate Three: Let $|\nu\rangle$ be the state of a system before the measurement $\{M_m\}$. The probability that result m occurs is given by

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

If result m occurs, then the state of the system after measurement is

$$\frac{M_m |\nu\rangle}{\sqrt{\langle \nu | M_m^\dagger M_m | \nu \rangle}}.$$

Notice that from Equation 2.16, the completeness equation, we have

$$\sum_m p(m) = \sum_m \langle \nu | M_m^\dagger M_m | \nu \rangle = \langle \nu | \left(\sum_m M_m^\dagger M_m \right) | \nu \rangle = \langle \nu | I | \nu \rangle = 1.$$

Thus, the sum of the probabilities of the measurement outcomes is equal to 1.

Example 3.5: Let $M_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$ and $M_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Observe that

$$\begin{aligned} \sum_{m=0}^1 M_m^\dagger M_m &= M_0^\dagger M_0 + M_1^\dagger M_1 \\ &= \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \right) + \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right) \\ &= \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I. \end{aligned}$$

Let $|\nu\rangle = \alpha|0\rangle + \beta|1\rangle$ be the state of the quantum system before measurement. The probability that measurement outcome 0 occurs is

$$\begin{aligned} p(0) &= \langle \nu | M_0^\dagger M_0 | \nu \rangle = \begin{bmatrix} \alpha^* & \beta^* \end{bmatrix} \left(\frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \alpha^* & \beta^* \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \frac{1}{2} (|\alpha|^2 + |\beta|^2) = \frac{1}{2} \end{aligned}$$

and the state of the system after measurement, if 0 occurs, is

$$\frac{M_0 |\nu\rangle}{\sqrt{\langle \nu | M_0^\dagger M_0 | \nu \rangle}} = \frac{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}}{\sqrt{\frac{1}{2}}} = \begin{bmatrix} \alpha \\ \beta i \end{bmatrix} = (\alpha|0\rangle + \beta i|1\rangle).$$

Similarly, the probability that measurement outcome 1 occurs is

$$p(1) = \langle \nu | M_1^\dagger M_1 | \nu \rangle = \frac{1}{2} (|\alpha|^2 + |\beta|^2) = \frac{1}{2}$$

and the state of the system after measurement, if 1 occurs, is

$$\frac{M_1 |\nu\rangle}{\sqrt{\langle \nu | M_1^\dagger M_1 | \nu \rangle}} = (\alpha|0\rangle - \beta|1\rangle).$$

$\{|0\rangle, |1\rangle\}$ is just one choice for basis states for a qubit. Given any orthonormal basis $\{|a\rangle, |b\rangle\}$ for a qubit system, by Theorem 3.4, it is possible to perform a measurement with respect to $\{|a\rangle, |b\rangle\}$.

Example 3.6: Let $|\nu\rangle = \alpha|0\rangle + \beta|1\rangle$ be an arbitrary state in the basis $\{|0\rangle, |1\rangle\}$. We measure $|\nu\rangle$ with respect to the basis states $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. In this case,

$$M_+ = |+\rangle\langle +| = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \text{ and}$$

$$M_- = |-\rangle\langle -| = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

The probability that measurement outcome $+$ occurs is

$$p(+) = \begin{bmatrix} \alpha^* & \beta^* \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \alpha^* & \beta^* \end{bmatrix} \begin{bmatrix} \alpha + \beta \\ \alpha + \beta \end{bmatrix}$$

$$= \frac{1}{2} (\alpha^* \alpha + \alpha^* \beta + \beta^* \alpha + \beta^* \beta) = \frac{1}{2} |\alpha + \beta|^2$$

and the state of the system after measurement, if $+$ occurs, is

$$\frac{M_+ |\nu\rangle}{\sqrt{\langle \nu | M_+^\dagger M_+ | \nu \rangle}} = \frac{\frac{1}{2}}{\sqrt{\frac{1}{2} |\alpha + \beta|^2}} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \frac{1}{\sqrt{2} |\alpha + \beta|} \begin{bmatrix} \alpha + \beta \\ \alpha + \beta \end{bmatrix}.$$

Similarly, the probability that measurement outcome $-$ occurs is

$$\begin{aligned}
 p(-) &= \begin{bmatrix} \alpha^* & \beta^* \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \alpha^* & \beta^* \end{bmatrix} \begin{bmatrix} \alpha - \beta \\ -\alpha + \beta \end{bmatrix} \\
 &= \frac{1}{2} (\alpha^* \alpha - \alpha^* \beta - \beta^* \alpha + \beta^* \beta) = \frac{1}{2} |\alpha - \beta|^2
 \end{aligned}$$

and the state of the system after measurement, if $-$ occurs, is

$$\frac{M_- |v\rangle}{\sqrt{\langle v | M_-^\dagger M_- | v \rangle}} = \frac{\frac{1}{2}}{\sqrt{\frac{1}{2} |\alpha - \beta|^2}} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \frac{1}{\sqrt{2} |\alpha - \beta|} \begin{bmatrix} \alpha - \beta \\ -\alpha + \beta \end{bmatrix}.$$

Often, a quantum system are found to be a composite of two or more quantum systems. The fourth postulate of quantum mechanics describes the state space of a composite quantum system. Underlying the fourth postulate is the notion of tensor product of vector spaces and vectors.

Postulate Four: If $H_1, H_2, \dots, H_n$ are state spaces of isolated quantum systems, then the composite quantum system, H , of these state spaces is the tensor product of their Hilbert spaces, i.e., $H = H_1 \otimes H_2 \otimes \dots \otimes H_n$. Moreover, if $|v_1\rangle, |v_2\rangle, \dots, |v_n\rangle$ are state vectors of the respective state spaces representing the n simultaneously held states of isolated systems, then $|v_1\rangle \otimes |v_2\rangle \otimes \dots \otimes |v_n\rangle$ is a state vector of H .

The converse to the second statement in Postulate Four, however, is not true. Not all state vectors in a composite space can be written as a tensor product of state vectors from the respective systems.

Example 3.7: If $H_1 = H_2 = \mathbb{C}^2$, the two qubit state $|\nu\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$, known as the Bell

state or EPR pair, cannot be written as a tensor of state vectors in $\mathbb{C}^2$.

Proof: Suppose there exist state vectors $|v_1\rangle = a|0\rangle + b|1\rangle$ and $|v_2\rangle = c|0\rangle + d|1\rangle$ such that

$|v_1\rangle \otimes |v_2\rangle = |\nu\rangle$. Notice that

$$\begin{aligned} |v_1\rangle \otimes |v_2\rangle &= (a|0\rangle + b|1\rangle) \otimes (c|0\rangle + d|1\rangle) \\ &= ac|00\rangle + ad|01\rangle + bc|10\rangle + bd|11\rangle. \end{aligned}$$

Computing coefficients, we see that $ad = 0$ and $bc = 0$, but $ac = \frac{1}{\sqrt{2}}$ and $bd = \frac{1}{\sqrt{2}}$ implies

that $a, b, c, d \neq 0$. This is a contradiction. $\square$

4 The Stern-Gerlach Experiment

In 1921 and 1922, Otto Stern and Walter Gerlach conducted an experiment in which silver atoms were beamed through a non-uniform magnetic field that deflects an atom in the direction of its magnetic moment. The field split the beams into parts corresponding to the magnetic moment. A depiction of the Stern-Gerlach experiment is shown in Figure 4.1. This experiment is an example of a physical interpretation of Postulates One and Three of quantum mechanics.

Figure 4.1 [Fran]. Stern-Gerlach apparatus.

Consider what happens as the beam of atoms is sent through the magnetic field. As the atom passes through the field, the magnetic moment of the atom rotates to align with the direction of the field. If the field is non-uniform, it forces the atoms to accelerate along the direction in which the field strength changes. This force is proportional to the projection of the magnetic moment in the direction of the field.

It was expected that the beam would split into $2l+1$ parts, where l is the total orbital angular momentum. However, when hydrogen atoms in their ground states ($l = 0$) were beamed through the non-uniform magnetic field, the beam did not remain unchanged, but rather, split into two. In this experiment, the beam of hydrogen atoms travels along the y -axis towards a magnetic field that deflects the atoms in the $\hat{z}$ direction, call this magnetic field SGZ. What emerged from the SGZ apparatus is two beams of particles, one in $+\hat{z}$ direction and the other in the $-\hat{z}$ direction, of equal intensity [Rae].

Since this was not the expected result, Stern and Gerlach observed the deflection of the atoms through a cascade of magnetic fields, shown in Figure 4.2. When the beams of particles exited the SGZ apparatus, the particles in the $-\hat{z}$ direction were blocked and the particles in the $+\hat{z}$ direction were sent through a SGX apparatus, a magnetic field that deflects the atoms in the $\hat{x}$ direction corresponding to the magnetic moment of the atom. Again two beams of particles emerged from the apparatus, one in $+\hat{x}$ direction and the other in the $-\hat{x}$ direction, of equal intensity.

Next, the particles in the $-\hat{x}$ direction were blocked and the particles in the $+\hat{x}$ direction were sent through another SGZ apparatus. They anticipated that only one beam would emerge this time because all the atoms with a $-\hat{z}$ magnetic moment were blocked from entering the second the magnetic field. To their dismay, two beams exited this SGZ apparatus.

Figure 4.2 [NiCh]. Three stage cascaded Stern-Gerlach measurements.

The cascaded Stern-Gerlach experiment provides direct evidence that the electron possesses an intrinsic angular momentum, called *spin*, independent of the rotation of the electron around the nucleus.

The qubit model can be used to explain the behavior of the atoms in the cascaded Stern-Gerlach experiment. The spin of the electron is given by $|\nu\rangle \in \mathbb{C}^2$, i.e., $\mathbb{C}^2$ is a Hilbert space describing this spin. The standard orthonormal basis for $\mathbb{C}^2$ is $\{|0\rangle, |1\rangle\}$. Another orthonormal basis for $\mathbb{C}^2$ is $\{|+\rangle, |-\rangle\}$ where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. We think of each Stern-Gerlach apparatus as a device that measures the spin of the atom in a given computational basis.

We will assume that there is only one beam in SGZ at any time and ignore the interactions between the atoms. Also, assuming we have an even distribution of spins in the $\hat{x}$ and $\hat{z}$ directions, the particles in the first beam that enter the SGZ apparatus have state $\nu = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}i|1\rangle$. The SGZ apparatus measures the spin in the computation basis

$\{|0\rangle, |1\rangle\}$, so the probability that the particle emerges in the $+\hat{z}$ is $p(0) = \frac{1}{2}$ and the probability that the particle emerges in the $-\hat{z}$ is $p(1) = \frac{1}{2}$.

Since the $|1\rangle$ particles are blocked, the state vector of the particles that enter the SGX apparatus is $|0\rangle$. The SGX apparatus then measures spin in the basis $\{|+\rangle, |-\rangle\}$.

We know, from Postulate Three of quantum mechanics, that the probability that $|0\rangle$ will have output $|+\rangle$ is

$$p(+)=|0\rangle^\dagger(|+\rangle\langle+|)|0\rangle=\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1 \\ 0\end{bmatrix}=\frac{1}{2},$$

and the state of the system after measurement, if $|+\rangle$ occurs, is

$$\frac{(|+\rangle\langle+|)|0\rangle}{\sqrt{\langle 0|(|+\rangle\langle+|)^\dagger(|+\rangle\langle+|)|0\rangle}}=\frac{1}{\sqrt{\frac{1}{2}}}\begin{bmatrix}\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1 \\ 0\end{bmatrix}=\begin{bmatrix}\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2}\end{bmatrix}=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)=|+\rangle.$$

The probability that $|0\rangle$ will have output $|-\rangle$ is

$$p(-)=|0\rangle^\dagger(|-\rangle\langle-|)|0\rangle=\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1 \\ 0\end{bmatrix}=\frac{1}{2},$$

and the state of the system after measurement, if $|-\rangle$, is

$$\frac{(|-\rangle\langle-|)|0\rangle}{\sqrt{\langle 0|(|-\rangle\langle-|)^\dagger(|-\rangle\langle-|)|0\rangle}}=\frac{1}{\sqrt{\frac{1}{2}}}\begin{bmatrix}\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1 \\ 0\end{bmatrix}=\begin{bmatrix}\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2}\end{bmatrix}=\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)=|-\rangle.$$

Next, the $|-\rangle$ particles are blocked, and the state vector of the particles that enter the SGZ apparatus is $|+\rangle$. The probability that $|+\rangle$ will have output $|0\rangle$ is

$$p(0) = |+\rangle^\dagger (|0\rangle\langle 0|) |+\rangle = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{1}{2},$$

and the state of the system after measurement, if $|0\rangle$, is

$$\frac{(|0\rangle\langle 0|) |+\rangle}{\sqrt{\langle + | (|0\rangle\langle 0|)^\dagger (|0\rangle\langle 0|) |+\rangle}} = \frac{1}{\sqrt{\frac{1}{2}}} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = |0\rangle.$$

The probability that $|+\rangle$ will have output $|1\rangle$ is

$$p(0) = |+\rangle^\dagger (|1\rangle\langle 1|) |+\rangle = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{1}{2},$$

and the state of the system after measurement, if $|1\rangle$, is

$$\frac{(|1\rangle\langle 1|) |+\rangle}{\sqrt{\langle + | (|1\rangle\langle 1|)^\dagger (|1\rangle\langle 1|) |+\rangle}} = \frac{1}{\sqrt{\frac{1}{2}}} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = |1\rangle.$$

This matches the observation that the $-\hat{z}$ (state $|1\rangle$ particles) return after being blocked from entry into the second SGZ apparatus.

5 Quantum Gates and Circuits

In this section, we will specify a precise model of quantum computation by giving a brief introduction to quantum gates and circuits. We will only study those that are of particular use in developing the quantum Fourier transform and factoring algorithm discussed in the later sections.

Quantum gates provide a way for describing how quantum machines manipulate quantum information. A quantum circuit is made of wires that carry information and quantum gates. In quantum circuits, wires (lines) represent qubits, '/' denotes a bundle of bits, time proceeds from left to right, and gates represent actions performed on qubits, that is, the evolution of qubit systems.

Example 5.1: Suppose we have a bundle, say n , of qubits initialized to the state $|0\rangle$. We want to perform a unitary transformation represented by the matrix

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

on each of the qubits. In this example, H is a quantum gate. The quantum circuit describing this process is depicted by the following diagram.

Reading this diagram from left to right, we see that a bundle of bits is being passed through the H gate. If the qubits are initialized to state $|0\rangle$, then the state that emerges from gate H is $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. By Postulate Four of quantum mechanics, the final state of the system is the tensor product of the individual states, i.e., the resulting system state is

$$\frac{1}{2^{n/2}}(|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \otimes \cdots \otimes (|0\rangle + |1\rangle) = \frac{1}{2^{n/2}} \bigotimes_{x=1}^n \sum_{y=0}^1 |y\rangle.$$

In the above equation, the vector $\frac{1}{2^{n/2}} \bigotimes_{x=1}^n \sum_{y=0}^1 |y\rangle$ is a vector in $\mathbb{C}^{2^n}$. It should be noted that a basis for the Hilbert space $\mathbb{C}^{2^n}$ is $\{|e_0\rangle, |e_1\rangle, \dots, |e_{2^n-1}\rangle\}$, where $|e_i\rangle$ is a vector with a 1 in the $(i+1)$ -th entry and zeros elsewhere. We represent an equal superposition of all basis states for $\mathbb{C}^{2^n}$ by $\sum_{i=0}^{2^n-1} a_i |e_i\rangle$, where the amplitudes a_i are complex numbers that satisfy the equation $\sum_{i=0}^{2^n-1} |a_i|^2 = 1$. For example, we may let $a_i = \frac{1}{2^{n/2}}$ for all i , since

$$\sum_{i=0}^{2^n-1} \left| \frac{1}{2^{n/2}} \right|^2 = \sum_{i=0}^{2^n-1} \frac{1}{2^{2n/2}} = \sum_{i=0}^{2^n-1} \frac{1}{2^n} = (2^n) \left(\frac{1}{2^n} \right) = 1.$$

So, the equal superposition of all states for $\mathbb{C}^{2^n}$ becomes

$$\sum_{i=0}^{2^n-1} \frac{1}{2^{n/2}} |e_i\rangle = \frac{1}{2^{n/2}} \bigotimes_{x=1}^n \sum_{y=0}^1 |y\rangle.$$

It follows that we can create an equal superposition of all basis states for $\mathbb{C}^{2^n}$ by applying Hadamard gates to n -qubits initialized to the state $|0\rangle$.

From the postulates of quantum mechanics, we know that only unitary operations are allowed on state systems. Thus, for any quantum circuit, there is a unitary matrix associated with it. Moreover, it can be shown that any unitary transformation is a feasible quantum circuit, although not all unitary transformations are necessarily practical or efficiently achievable.

Consider the following example of a quantum circuit.

Example 5.2: Suppose a unitary transformation acts on the basis elements of $(\mathbb{C}^2)^{\otimes 2}$ as follows:

$$|00\rangle \rightarrow \frac{1}{\sqrt{2}}(|00\rangle + |01\rangle)$$

$$|01\rangle \rightarrow \frac{1}{\sqrt{2}}(|00\rangle - |01\rangle)$$

$$|10\rangle \rightarrow |10\rangle$$

$$|11\rangle \rightarrow |11\rangle.$$

Relative to the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, this circuit is given by the unitary matrix

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where each row corresponds to the input basis vectors and each column corresponds to the output basis vectors. If we apply this circuit to a state system in a superposition of states, say $|\nu\rangle = a_0|00\rangle + a_1|01\rangle + a_2|10\rangle + a_3|11\rangle$, the resulting superposition of states is given by multiplication of the vector $|\nu\rangle$ with the matrix U , i.e., we get

$$\begin{aligned}
 U|v\rangle &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \\
 &= \begin{bmatrix} \frac{1}{\sqrt{2}}a_0 + \frac{1}{\sqrt{2}}a_1 \\ \frac{1}{\sqrt{2}}a_0 - \frac{1}{\sqrt{2}}a_1 \\ a_2 \\ a_3 \end{bmatrix} \\
 U|v\rangle &= \frac{1}{\sqrt{2}}(a_0 + a_1)|00\rangle + \frac{1}{\sqrt{2}}(a_0 - a_1)|01\rangle + a_2|10\rangle + a_3|11\rangle.
 \end{aligned}$$

The matrix U is an example of a two qubit circuit. It is represented by

The simplest quantum gates are operations on single qubits. Single qubit gates are described by 2×2 unitary matrices. It will be of use to list them here with the schematic symbols that are often used to denote them. The notation $E(x)$ is used to denote $e^{2\pi i x}$.

Hadamard	$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$	
----------	--	--

R_t	$\begin{bmatrix} 1 & 0 \\ 0 & E(2^{-t}) \end{bmatrix}$	
-------	--	--

Pauli-X	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	
---------	--	--

Although all these gates are equally realistic, the Hadamard and R_z gates are especially useful in creating a circuit to perform the quantum Fourier transform. The Pauli- X matrix is often referred to as the *bit flip* matrix since it takes $|0\rangle$ to $|1\rangle$ and $|1\rangle$ to $|0\rangle$.

Another important gate is the control-NOT or CNOT gate. The CNOT gate has two input qubits, a control qubit and a target qubit. This gate negates the target qubit if and only if the control bit is set to 1. The matrix representation for the CNOT gate is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

It is symbolized by the circuit

Here, the top line is the control qubit and the bottom line is the target qubit.

A more general control operation is the control- U operation, where U is any fixed single qubit unitary operation. Similar to the CNOT operation, the control- U operation also contains a control qubit and a target qubit. However, in this operation, setting the control bit to 1 applies the U -operator to the target qubit. This circuit is depicted by

Actually, Example 5.2 is a control- H surrounded by bit flips. So, it can be represented by the circuit

Example 5.3: Let $|1\rangle$ be the control qubit, $|v\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ be the target qubit and

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

Since the control bit is set to 1, the matrix U is applied to the target qubit. Notice that this unitary transformation preserves the $|0\rangle$ qubit and sends the $|1\rangle$ qubit to its negative, relative to the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. Therefore, $|v\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. (In fact, U is control- R_1 .)

Control operations can also be used to construct swap operations. A swap operation interchanges the states of two qubits. It sends $|x\rangle \otimes |y\rangle$ to $|y\rangle \otimes |x\rangle$. This is performed using three control-NOT gates. The matrix representation of the swap operation is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and the circuit representation is

For the purposes of this paper, we will assume that we can only use the gates given in Table 5.4 in a quantum computer.

List of Allowable Gates	
1.	Hadamard gate H
2.	Bit flip (Pauli) gate, X
3.	Control-NOT gate, CNOT
4.	R_k and R_k^{-1} gates for all positive integers k
5.	Control- R_k and control- R_k^{-1} gates for all positive integers k

Table 5.4. List of Allowable Gates.

Note that these five types of gates, like all quantum gates, are all reversible gates, that is, the input can always be recovered from the output of these operations. Thus, no information is ever erased.

6 Quantum Fourier Transform

This section is devoted to the study of the quantum Fourier transform (QFT). It is the foundation of the quantum factoring algorithm. We will show that the QFT is both feasible and efficient in a quantum computer by constructing a quantum circuit to compute the QFT.

Let $\{|0\rangle, |1\rangle, \dots, |N-1\rangle\}$ be an orthonormal basis of the Hilbert space $\mathbb{C}^N$ with inner product $\langle x|y\rangle = |x\rangle^\dagger |y\rangle$ and norm $\|x\| = \sqrt{\langle x|x\rangle}$. The discrete Fourier transform is the transformation $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$ whose action on the given basis is defined by

$$|x\rangle \mapsto \frac{1}{\sqrt{N}} \sum_{y=0}^{N-1} E(xy/N) |y\rangle \quad (6.1)$$

where $E(\mu) = e^{2\pi i \mu}$. It follows from the definition of basis that for any arbitrary vector $|v\rangle$, we have

$$|v\rangle = \sum_{x=0}^{N-1} a_x |x\rangle.$$

Hence, F sends

$$\begin{aligned} |v\rangle &\mapsto \sum_{x=0}^{N-1} a_x \left(\frac{1}{\sqrt{N}} \sum_{y=0}^{N-1} E(xy/N) |y\rangle \right) \\ &= \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} a_x \left(\sum_{y=0}^{N-1} E(xy/N) |y\rangle \right) \\ &= \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} \left(\sum_{y=0}^{N-1} a_x E(xy/N) \right) |y\rangle \\ &= \sum_{y=0}^{N-1} b_y |y\rangle \end{aligned}$$

where $b_y = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} a_x E(xy/N)$.

Let $N = 2^n$, where n is a natural number. We define the qubit representation of a basis element, $|x\rangle$, of $\mathbb{C}^{2^n}$ to be $|x\rangle = |x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_n\rangle$, where $|x_i\rangle$ is either $|0\rangle$ or $|1\rangle$ in $\mathbb{C}^2$, and $x_1, x_2, \dots, x_n$ are binary digits of x , i.e.,

$$x = x_1 2^{n-1} + x_2 2^{n-2} + \cdots + x_n 2^0 = \sum_{i=1}^n x_i 2^{n-i}. \text{ In other words, we think of } \mathbb{C}^{2^n} \text{ as } (\mathbb{C}^2)^{\otimes n}.$$

The table below gives the binary and qubit representations of the basis elements of $\mathbb{C}^8$.

Index	Binary Representation	Qubit Representation
0	0	$ 000\rangle$
1	1	$ 001\rangle$
2	10	$ 010\rangle$
3	11	$ 011\rangle$
4	100	$ 100\rangle$
5	101	$ 101\rangle$
6	110	$ 110\rangle$
7	111	$ 111\rangle$

Table 6.2. Qubit representation of basis for $\mathbb{C}^8$.

Example 6.3: Let $|x\rangle = |00\dots 0\rangle$ be an n -qubit state in $(\mathbb{C}^2)^{\otimes n}$. The discrete Fourier transform on $|x\rangle$ is

$$\begin{aligned}
|x\rangle &= |00\dots 0\rangle \rightarrow \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n-1} E(xy/2^n) |y\rangle \\
&= \frac{1}{\sqrt{2^n}} E(0) \sum_{y=0}^{2^n-1} |y\rangle \\
&= \frac{1}{\sqrt{2^n}} (1) \sum_{y=0}^{2^n-1} |y\rangle \\
&= \frac{1}{\sqrt{2^n}} (|0\rangle + |1\rangle + \dots + |N-1\rangle)
\end{aligned}$$

since $x = \sum_{m=1}^n x_m 2^m = 0 \cdot 2^1 + 0 \cdot 2^2 + \dots + 0 \cdot 2^n = 0$.

Lemma 6.4: Let $\omega \in \mathbb{Z}$. Then

$$\sum_{a=0}^{N-1} E\left(\frac{\omega a}{N}\right) = \begin{cases} N & \text{if } \omega \equiv 0 \pmod{N} \\ 0 & \text{if } \omega \not\equiv 0 \pmod{N}. \end{cases}$$

Proof: When $\omega \equiv 0 \pmod{N}$,

$$\sum_{a=0}^{N-1} E\left(\frac{\omega a}{N}\right) = \sum_{a=0}^{N-1} 1 = N.$$

For $\omega \not\equiv 0 \pmod{N}$, let $\zeta = E\left(\frac{\omega}{N}\right)$. So, $\zeta^N = 1$, implying that

$$\zeta^N - 1 = (\zeta - 1)(\zeta^{N-1} + \dots + \zeta + 1) = 0.$$

Since $\omega \not\equiv 0 \pmod{N}$, $(\zeta - 1) \neq 0$. Therefore,

$$\sum_{a=0}^{N-1} E\left(\frac{\omega a}{N}\right) = \sum_{a=0}^{N-1} \zeta^a = \zeta^{N-1} + \dots + \zeta + 1 = 0. \quad \square$$

Proposition 6.5: The transformation F described in Equation 6.1 is a unitary operator.

Proof: Let $|x\rangle$ and $|y\rangle$ be two distinct orthonormal basis elements of $\mathbb{C}^N$. By Theorem

2.5, it suffices to show that $\langle x|F^\dagger F|x\rangle = 1$ and $\langle x|F^\dagger F|y\rangle = 0$. Note that if $|x\rangle \neq |y\rangle$,

then $\langle x|x\rangle = 1$ and $\langle x|y\rangle = 0$, since $|x\rangle$ and $|y\rangle$ are elements of an orthonormal basis.

Also, by Lemma 6.4, it follows that

$$\begin{aligned}
 \langle x|F^\dagger F|x\rangle &= \langle F|x\rangle, F|x\rangle\rangle \\
 &= \left\langle \frac{1}{\sqrt{N}} \sum_{a=0}^{N-1} E(xa/N)|a\rangle, \frac{1}{\sqrt{N}} \sum_{b=0}^{N-1} E(xb/N)|b\rangle \right\rangle \\
 &= \left[\frac{1}{\sqrt{N}} \sum_{a=0}^{N-1} E(xa/N)|a\rangle \right]^\dagger \left[\frac{1}{\sqrt{N}} \sum_{b=0}^{N-1} E(xb/N)|b\rangle \right] \\
 &= \frac{1}{N} \left[\sum_{a=0}^{N-1} E(xa/N)^* |a\rangle^\dagger \right] \left[\sum_{b=0}^{N-1} E(xb/N)|b\rangle \right] \\
 &= \frac{1}{N} \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} E(xa/N)^* E(xb/N) \langle a|b\rangle \\
 &= \frac{1}{N} \sum_{a=0}^{N-1} E(xa/N)^* E(xa/N) \\
 &= \frac{1}{N} (N) \\
 &= 1,
 \end{aligned}$$

and

$$\begin{aligned}
\langle x | F^\dagger F | y \rangle &= \langle F | x \rangle, F | y \rangle \\
&= \left\langle \frac{1}{\sqrt{N}} \sum_{a=0}^{N-1} E(xa/N) | a \rangle, \frac{1}{\sqrt{N}} \sum_{b=0}^{N-1} E(yb/N) | b \rangle \right\rangle \\
&= \frac{1}{N} \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} E(xa/N)^* E(yb/N) \langle a | b \rangle \\
&= \frac{1}{N} \sum_{a=0}^{N-1} E(xa/N)^* E(ya/N) \\
&= \frac{1}{N} \sum_{a=0}^{N-1} E((y-x)a/N) \\
&= \frac{1}{N} (0) \\
&= 0. \quad \square
\end{aligned}$$

Since F is a unitary operator, it follows from Postulate Two of quantum mechanics that the quantum Fourier transform can be realized, i.e., a quantum circuit can be constructed to implement the quantum Fourier transform. However, in order to construct an efficient circuit using the gates listed in Section 5, it is necessary to derive a different representation of the quantum Fourier transform.

Notice that $\frac{y}{2^n} = 2^{-n} \sum_{m=1}^n y_m 2^{n-m} = \sum_{m=1}^n y_m 2^{-m}$. Therefore, using the qubit representation of a state vector in $(\mathbb{C}^2)^{\otimes n}$, the discrete Fourier transform becomes

$$\begin{aligned}
|x\rangle &\rightarrow \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n-1} E(xy/2^n) |y\rangle \\
&= \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n-1} E\left(x \sum_{m=1}^n y_m 2^{-m}\right) |y\rangle \\
&= \frac{1}{\sqrt{2^n}} \sum_{y_1=0}^1 \sum_{y_2=0}^1 \cdots \sum_{y_n=0}^1 E\left(x \sum_{m=1}^n y_m 2^{-m}\right) |y_1 y_2 \cdots y_n\rangle \\
&= \frac{1}{\sqrt{2^n}} \sum_{y_1=0}^1 \sum_{y_2=0}^1 \cdots \sum_{y_n=0}^1 \prod_{m=1}^n E(xy_m 2^{-m}) |y_1 y_2 \cdots y_n\rangle \\
&= \frac{1}{\sqrt{2^n}} \sum_{y_1=0}^1 \sum_{y_2=0}^1 \cdots \sum_{y_n=0}^1 \left[\bigotimes_{m=1}^n E(xy_m 2^{-m}) |y_m\rangle \right] \\
&= \frac{1}{\sqrt{2^n}} \bigotimes_{m=1}^n \left[\sum_{y_m=0}^1 E(xy_m 2^{-m}) |y_m\rangle \right] \\
&= \frac{1}{\sqrt{2^n}} \bigotimes_{m=1}^n [|0\rangle + E(x 2^{-m}) |1\rangle].
\end{aligned}$$

The expression

$$\frac{1}{\sqrt{2^n}} \bigotimes_{m=1}^n [|0\rangle + E(x 2^{-m}) |1\rangle]$$

is called the product representation of the quantum Fourier transform. Note that

$E(x 2^{-m})$ is $E(\text{last } m \text{ binary digits of } x)$ because we can write x in binary decimal

notation as $x_0 x_1 x_2 \dots x_n = (x_0/2) + (x_1/2^2) + (x_2/2^3) + \cdots (x_n/2^{n+1})$. Note also that

$E(a) = 1$ for all $a \in \mathbb{Z}$.

Example 6.5: The product representation of the quantum Fourier transform of $|x\rangle = |6\rangle$

is

$$\begin{aligned}
 |110\rangle &\mapsto \frac{(|0\rangle + E(11.0)|1\rangle) \otimes (|0\rangle + E(1.10)|1\rangle) \otimes (|0\rangle + E(.110)|1\rangle)}{2^3} \\
 &= \frac{(|0\rangle + E(.0)|1\rangle) \otimes (|0\rangle + E(.10)|1\rangle) \otimes (|0\rangle + E(.110)|1\rangle)}{2^3}
 \end{aligned}$$

Using the product representation, the following three-qubit circuit can be constructed for the quantum Fourier transform.

Figure 6.6. Three-qubit QFT circuit.

Here, H represents the Hadamard transformation, and the other gates are control- R_k gates, along with a swap operation at the end.

From the pictured circuit, we can see that when the states $|x_1\rangle$, $|x_2\rangle$, and $|x_3\rangle$ enter the circuit, the Hadamard gate is applied to the first qubit, producing the state

$$H|x_1\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} |x_1\rangle = \frac{1}{\sqrt{2}} \sum_{y=0}^1 (-1)^{xy} |y\rangle = \frac{1}{\sqrt{2}} (|0\rangle + E(0.x_1)|1\rangle)$$

in the first line. After this, the R_2 gate is applied to the new state under the control of the second qubit. In this case, the R_2 gate leaves the $|0\rangle$ qubit alone and multiplies the $|1\rangle$

qubit by a factor of $E(2^{-2})$ if the second qubit $|x_2\rangle$ is set to $|1\rangle$. Thus, applying the R_2 gate to the new state yields

$$\begin{aligned}
 & R_2 \left[\frac{1}{\sqrt{2}} (|0\rangle + E(0.x_1)|1\rangle) \right] \otimes |x_2\rangle \\
 &= \frac{1}{\sqrt{2}} (|0\rangle + E(0.x_1)E(x_2/2^2)|1\rangle) \otimes |x_2\rangle \\
 &= \frac{1}{\sqrt{2}} (|0\rangle + E(x_1/2 + x_2/2^2)|1\rangle) \otimes |x_2\rangle \\
 &= \frac{1}{\sqrt{2}} (|0\rangle + E(0.x_1x_2)|1\rangle) \otimes |x_2\rangle \\
 &= \frac{1}{\sqrt{2}} (|0\rangle + E(0.x_1x_2)|1\rangle) \otimes |x_2\rangle
 \end{aligned}$$

in the first two lines. Now, the R_3 gate is applied to this state under the control of the third qubit. The R_3 gate leaves the $|0\rangle$ qubit unchanged and multiplies the $|1\rangle$ qubit by a factor of $E(2^{-3})$ if the third qubit $|x_3\rangle$ is set to $|1\rangle$. So the state of the system becomes

$$\begin{aligned}
 & R_3 \left[\frac{1}{\sqrt{2}} (|0\rangle + E(0.x_1x_2)|1\rangle) \right] \otimes |x_2\rangle \otimes |x_3\rangle \\
 &= R_3 \left[\frac{1}{\sqrt{2}} (|0\rangle + E(0.x_1)E(x_2/2^2)|1\rangle) \otimes |x_2\rangle \otimes |x_3\rangle \right] \\
 &= \frac{1}{\sqrt{2}} (|0\rangle + E(x_1/2 + x_2/2^2 + x_3/2^3)|1\rangle) \otimes |x_2\rangle \otimes |x_3\rangle \\
 &= \frac{1}{\sqrt{2}} (|0\rangle + E(0.x_1x_2x_3)|1\rangle) \otimes |x_2\rangle \otimes |x_3\rangle \\
 &= \frac{1}{\sqrt{2}} (|0\rangle + E(0.x_1x_2x_3)|1\rangle) \otimes |x_2\rangle \otimes |x_3\rangle.
 \end{aligned}$$

Next, perform a similar procedure on the second qubit $|x_2\rangle$. The Hadamard and

R_2 gates produce the states

$$\frac{1}{2} (|0\rangle + E(0.x_1x_2x_3)|1\rangle) \otimes (|0\rangle + E(0.x_2)|1\rangle) \otimes |x_3\rangle$$

and

$$\frac{1}{2}(|0\rangle + E(0.x_1x_2x_3)|1\rangle) \otimes (|0\rangle + E(0.x_2x_3)|1\rangle) \otimes |x_3\rangle$$

respectively.

Finally, the single Hadamard gate on the last qubit transforms $|x_3\rangle$ into the state

$$\frac{1}{\sqrt{2}}(|0\rangle + E(0.x_3)|1\rangle).$$

At that point, the state of the system is

$$\frac{(|0\rangle + E(0.x_1x_2x_3)|1\rangle) \otimes (|0\rangle + E(0.x_2x_3)|1\rangle) \otimes (|0\rangle + E(0.x_3)|1\rangle)}{2^{3/2}}.$$

Swap operations are then used to reverse the order of the qubits to obtain

$$\frac{(|0\rangle + E(0.x_3)|1\rangle) \otimes (|0\rangle + E(0.x_2x_3)|1\rangle) \otimes (|0\rangle + E(0.x_1x_2x_3)|1\rangle)}{2^{3/2}}.$$

Analogously, the circuit in Figure 6.7 provides an $O(n^2)$ algorithm for the QFT on n -qubits.

Figure 6.7 [NiCh]. An n -qubit QFT circuit.

As the first qubit enters this circuit, the Hadamard gate is applied to it, producing the state $\frac{1}{\sqrt{2}}(|0\rangle + E(0.x_1)|1\rangle) \otimes |x_2\rangle \otimes |x_3\rangle \otimes \dots \otimes |x_n\rangle$. Next, the R_2 gate applied to the new state under the control of $|x_2\rangle$. This yields the state

$$\frac{1}{\sqrt{2}}(|0\rangle + E(0.x_1x_2)|1\rangle) \otimes |x_2\rangle \otimes |x_3\rangle \otimes \dots \otimes |x_n\rangle.$$

Then, the controlled R_3 gate is applied to this state, resulting in the state

$$\frac{1}{\sqrt{2}}(|0\rangle + E(0.x_1x_2x_3)|1\rangle) \otimes |x_2\rangle \otimes |x_3\rangle \otimes \dots \otimes |x_n\rangle.$$

Notice that each R_k gate contributes a factor of $E(x_k/2^k)$ to the coefficient of $|1\rangle$ in the first qubit. Continuing, we see that the final state at the end of the first line in the circuit is $\frac{1}{\sqrt{2}}(|0\rangle + E(0.x_1x_2x_3\dots x_n)|1\rangle) \otimes |x_2\rangle \otimes |x_3\rangle \otimes \dots \otimes |x_n\rangle$.

In the next line of the circuit, $|x_2\rangle$ goes through a similar procedure. First, it is passed through the Hadamard gate, yielding a state of

$$\frac{1}{2}(|0\rangle + E(0.x_1x_2\dots x_n)|1\rangle) \otimes (|0\rangle + E(0.x_2)|1\rangle) \otimes |x_3\rangle \otimes |x_4\rangle \otimes \dots \otimes |x_n\rangle.$$

The series of controlled R_k produces the state

$$\frac{1}{2}(|0\rangle + E(0.x_1x_2\dots x_n)|1\rangle) \otimes (|0\rangle + E(0.x_2x_3\dots x_n)|1\rangle) \otimes |x_3\rangle \otimes |x_4\rangle \otimes \dots \otimes |x_n\rangle.$$

In the same way, applying the Hadamard gate to the i th qubit $|x_i\rangle$ gives rise to the state

$$\frac{1}{2^{i/2}}(|0\rangle + E(0.x_1x_2\dots x_n)|1\rangle) \otimes (|0\rangle + E(0.x_2x_3\dots x_n)|1\rangle) \otimes \dots \otimes (|0\rangle + E(0.x_i)|1\rangle) \otimes |x_i\rangle \otimes |x_{i+1}\rangle \otimes \dots \otimes |x_n\rangle$$

and the series of controlled R_k yields

$$\frac{1}{2^{i/2}}(|0\rangle + E(0.x_1x_2\dots x_n)|1\rangle) \otimes (|0\rangle + E(0.x_2x_3\dots x_n)|1\rangle) \otimes \dots \otimes (|0\rangle + E(0.x_ix_{i+1}\dots x_n)|1\rangle) \otimes |x_{i+1}\rangle \otimes |x_{i+2}\rangle \otimes \dots \otimes |x_n\rangle$$

at the end of the i th line. Thus, the final state of the system is

$$\frac{1}{2^{n/2}}(|0\rangle + E(0.x_1x_2\dots x_n)|1\rangle) \otimes (|0\rangle + E(0.x_2x_3\dots x_n)|1\rangle) \otimes \dots \otimes (|0\rangle + E(0.x_n)|1\rangle).$$

The last part of this circuit, which is not shown in Figure 6.6, swaps the states of the qubits. Since swap operations send $|x\rangle \otimes |y\rangle$ to $|y\rangle \otimes |x\rangle$, we may swap the first and last tensor factors, second and next to last, and so on. This produces the final state

$$\frac{1}{\sqrt{2^n}} \bigotimes_{n=1}^n [|0\rangle + E(x2^{-n})|1\rangle].$$

Theorem 6.8: The quantum Fourier transform on $\mathbb{C}^N$, $N = 2^L$, can be done with $O(L^2)$ quantum gates from the list in Table 5.4.

Proof: From Figure 6.7, it can easily be seen that the first part of the quantum Fourier circuit uses $L + (L-1) + (L-2) + \dots + 2 + 1 = L(L+1)/2$ gates. Then swap operations are performed to reverse the order of the qubits. At most $L/2$ swaps are required, and each swap can be performed using three control-NOT gates. Thus, the quantum Fourier transform can be done with $O(L^2)$ gates from the list in Table 5.4. $\square$

7 Number Theory

The quantum factoring algorithm uses several concepts from number theory. Unless otherwise cited, most of the theorems presented in this section can be found in Ireland and Rosen [IrRo].

Two well-known results from number theory are the Möbius Inversion Theorem and Fermat's Little Theorem. The next three lemmas are used to prove the Mobius Inversion Theorem.

We consider functions on positive integers. The Dirichlet product of functions f and g is $f * g(n) = \sum f(a)g(b)$ where the sum is over all pairs of positive integers (a, b) such that $ab = n$.

Lemma 7.1: The Dirichlet product is associative and commutative.

Proof: By definition,

$$f * (g * h)(n) = \sum f(a)g(p)h(q) = (f * g) * h(n)$$

where the sum is over all 3-tuples such that (a, p, q) such that $apq = n$. We also have

$$f * g(n) = \sum f(a)g(b) = \sum g(b)f(a) = g * f(n).$$

Therefore, the Dirichlet product is both associative and commutative. $\square$

Now, we define a function μ on $\mathbb{Z}^+$ by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n \text{ is not square free} \\ (-1)^k & \text{if } n = p_1 p_2 \dots p_k \text{ for distinct positive primes } p_i. \end{cases}$$

Lemma 7.2: If $n > 1$, then $\sum_{d|n} \mu(d) = 0$.

Proof: For $n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$, where $p_1, p_2, \dots, p_k$ are distinct primes, we have

$$\sum_{d|n} \mu(d) = \sum_{(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k)} \mu(p_1^{\varepsilon_1} p_2^{\varepsilon_2} \dots p_k^{\varepsilon_k})$$

such that ε_i is either 0 or 1. So,

$$\begin{aligned} \sum_{d|n} \mu(d) &= \binom{k}{0} + (-1) \binom{k}{1} + (-1)^2 \binom{k}{2} + \dots + (-1)^k \binom{k}{k} \\ &= (1-1)^k = 0. \quad \square \end{aligned}$$

The set of all units of $\mathbb{Z}_n$ forms a group under multiplication. This group is denoted by $\mathbb{Z}_n^*$. Recall that the Euler phi-function $\phi(n)$ is the number of positive integers less than or equal to n and relatively prime to n , which is the same as the number of elements of $\mathbb{Z}_n$ that are not divisors of 0. The Euler-phi function has the following properties [IrRo]:

(i) $\phi(mn) = \phi(m)\phi(n)$ if m is relatively prime to n .

(ii) $\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$ where $n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$ and $p_1, p_2, \dots, p_k$ are distinct primes.

(iii) $\sum_{d|n} \phi(d) = n$.

Now, we define two more functions Y, I on $\mathbb{Z}^+$ by $Y(1) = 1$ and $Y(n) = 0$ for $n > 1$, and $I(n) = 1$ for all $n \in \mathbb{Z}^+$.

Lemma 7.3: $I * \mu = \mu * I = \Upsilon$.

Proof: When $n=1$, $\mu * I(1) = \mu(1)I(1) = 1 \cdot 1 = 1$. If $n > 1$, then

$$\mu * I(n) = \sum_{d|n} \mu(d) = 0. \quad \mu * I = \Upsilon \text{ follows by commutativity. } \square$$

Möbius Inversion Theorem: Let $F(n) = \sum_{d|n} f(d)$. Then $f(n) = \sum_{d|n} \mu(d)F(n/d)$.

Proof: $F = f * I$ since $I(n) = 1$ for all $n \in \mathbb{Z}^+$. Hence,

$$F * \mu = (f * I) * \mu = f * (I * \mu) = f * \Upsilon = f.$$

Therefore, $f(n) = F * \mu(n) = \sum_{d|n} \mu(d)F(n/d)$. $\square$

Fermat's Little Theorem: If p is prime and $p \nmid x$, then $x^{p-1} \equiv 1 \pmod{p}$.

Proof: The units of $\mathbb{Z}_p$ form a group of order $p-1$. Since $p \nmid x$, x is a unit of $\mathbb{Z}_p$.

Therefore, the order of x divides the order of $\mathbb{Z}_p^*$, and $x^{p-1} \equiv 1 \pmod{p}$. $\square$

Next, we will use several lemmas to show that $\mathbb{Z}_p^*$ is a cyclic group. Once this is proved, we can show that $\mathbb{Z}_{p^k}^*$ is also a cyclic group.

Lemma 7.4: If F is a field and $f(x) \in F[x]$, then $(x-a) \mid f(x)$ if and only if

$$f(a) = 0.$$

Proof: Suppose $(x-a) \mid f(x)$. We can write $f(x) = (x-a)g(x)$ for some

$g(x) \in F[x]$ where the degree of $g(x)$ is less than the degree of $f(x)$. Hence,

$$f(a) = (a-a)g(a) = 0 \cdot g(a) = 0.$$

On the other hand, if $f(a) = 0$, then a is a zero of $f(x)$, i.e., $(x-a)$ is a factor of $f(x)$. Therefore, $(x-a) \mid f(x)$. $\square$

Lemma 7.5: Let $f(x) \in F[x]$ where F is a field and $\deg f(x) = n$. Then f has at most n distinct roots.

Proof: We will prove this by induction on n . When $n = 1$, it is trivial that $\deg f(x) = 1$ implies that f has at most one distinct root. Assume this is true for $n = k$. If $f(x)$ has no roots in F , then we are done. If δ is a root in F , then $f(x) = q(x)(x - \delta)$ by Lemma 7.4. So, $q(x)$ has degree $n - 1$. If $\gamma \neq \delta$ is another root of $f(x)$, then $0 = f(\gamma) = q(\gamma)(\gamma - \delta)$. Since $(\gamma - \delta) \neq 0$, it follows that $q(\gamma) = 0$. By induction hypothesis, $q(x)$ has at most $n - 1$ distinct roots. Therefore, $f(x)$ has at most n distinct roots. $\square$

Lemma 7.6: Let p be a prime number. Then $x^{p-1} - 1 = (x-1)(x-2)\dots(x-(p-1))$.

Proof: For any $a \in \mathbb{Z}_p^*$, $(x-a) \mid (x^{p-1} - 1)$ by Lemma 7.4 and Fermat's Little Theorem.

So, $(x-1)(x-2)\dots(x-(p-1)) \mid (x^{p-1} - 1)$. By comparing degrees and leading terms, we see that $(x-1)(x-2)\dots(x-(p-1))$ must equal $(x^{p-1} - 1)$. $\square$

Lemma 7.7: If $d \mid p-1$, then $x^d \equiv 1 \pmod{p}$ has exactly d solutions.

Proof: Let $dd' = p-1$, then

$$\frac{x^{p-1}-1}{x^d-1} = \frac{(x^d)^{d'}-1}{x^d-1} = (x^d)^{d'-1} + (x^d)^{d'-2} + \dots + x^d + 1 = g(x).$$

So, $g(x)(x^d-1) = x^{p-1}-1$. By Lemma 7.6, $x^{p-1}-1 = (x-1)(x-2)\dots(x-p+1)$, so

$(x^{p-1}-1)$ has $p-1$ distinct roots. Notice that $g(x)$ has at most

$p-1-d = dd'-d = d(d'-1)$ distinct roots. If (x^d-1) has fewer than d distinct roots,

then $g(x)(x^d-1)$ would have $k < p-1$ distinct roots. Therefore, since

$g(x)(x^d-1) = x^{p-1}-1$, we may conclude that x^d-1 must have exactly d roots. $\square$

Theorem 7.8: Let p be a prime number. Then $\mathbb{Z}_p^*$ is a cyclic group.

Proof: Let $d \in \mathbb{Z}^+$ such that $d|p-1$ and let $\psi(d)$ be the number of elements of $\mathbb{Z}_p^*$ of order d . By Lemma 7.7, the elements $x \in \mathbb{Z}_p^*$ that satisfy $x^d \equiv 1 \pmod{p}$ form a group of

order d . Thus, $\sum_{c|d} \psi(c) = d$. By applying the Möbius Inversion Theorem, we obtain

$\psi(d) = \sum_{c|d} \mu(c)(d/c)$. Since $d = \sum_{c|d} \phi(c)$, the Möbius Inversion Theorem also

implies that $\phi(d) = \sum_{c|d} \mu(c)(d/c)$, hence $\psi(d) = \phi(d)$. In particular,

$\psi(p-1) = \phi(p-1) > 1$ if $p > 2$. The case when $p = 2$ is trivial. So, there exist $\phi(p-1)$

elements of order $p-1$ in $\mathbb{Z}_p^*$. $\square$

Using Theorem 7.8 and the following three lemmas, we can show that $\mathbb{Z}_{p^\alpha}^*$ is also a cyclic group.

Lemma 7.9: If $\alpha \geq 1$ and $a \equiv b \pmod{p^\alpha}$, then $a^p \equiv b^p \pmod{p^{\alpha+1}}$.

Proof: Given $a \equiv b \pmod{p^\alpha}$, we have $a = b + cp^\alpha$ for some integer c . It follows that

$$\begin{aligned} a^p &= (b + cp^\alpha)^p \\ &= b^p + \binom{p}{1} b^{p-1} cp^\alpha + \binom{p}{2} b^{p-2} (cp^\alpha)^2 + \cdots + \binom{p}{p} (cp^\alpha)^p \\ &= b^p + \binom{p}{1} b^{p-1} cp^\alpha + G. \end{aligned}$$

G is divisible by $p^{\alpha+1}$ since each term is divisible by $p^{2\alpha}$ and $\alpha \geq 1$. Moreover,

$\binom{p}{1} b^{p-1} cp^\alpha$ is clearly divisible by $p^{\alpha+1}$. Therefore, $a^p \equiv b^p \pmod{p^{\alpha+1}}$. $\square$

Lemma 7.10: If $\alpha \geq 2$ and $p \neq 2$, then $(1 + ap)^{p^{\alpha-2}} \equiv 1 + ap^{\alpha-1} \pmod{p^\alpha}$ for all $\alpha \in \mathbb{Z}$.

Proof: We will prove this by induction on α . When $\alpha = 2$, we have

$$(1 + ap)^{p^{2-2}} = (1 + ap)^{p^0} = 1 + ap \pmod{p^2}.$$

Suppose the lemma holds for $\alpha = n \geq 2$. By Lemma 7.9, we have

$$\begin{aligned} (1 + ap)^{p^{n-2}} &\equiv 1 + ap^{n-1} \pmod{p^n} \\ \left[(1 + ap)^{p^{n-2}} \right]^p &\equiv (1 + ap^{n-1})^p \pmod{p^{n+1}} \\ (1 + ap)^{p^{n-1}} &\equiv (1 + ap^{n-1})^p \pmod{p^{n+1}}. \end{aligned}$$

Also,

$$\begin{aligned} (1+ap^{n-1})^p &= 1 + \binom{p}{1}ap^{n-1} + \binom{p}{2}(ap^{n-1})^2 + \cdots + \binom{p}{p}(ap^{n-1})^p \\ &= 1 + \binom{p}{1}ap^{n-1} + G. \end{aligned}$$

where G is a sum of $p-2$ terms. Since p divides $\binom{p}{k}$ for $0 < k < p$ and $p^{2(n-1)}$ divides

$(ap^{n-1})^k$ for $k = 2, 3, \dots, p$, it follows that $p^{1+2(n-1)}$ divides every term in G , except

possibly the last term $\binom{p}{p}(ap^{n-1})^p$. Since $n \geq 2$ and $p \geq 3$, we have $1+2(n-1) \geq n+1$

and $p(n-1) \geq n+1$. Hence, G is divisible by p^{n+1} . Also, $\binom{p}{1}ap^{n-1} = pap^{n-1} = ap^n$.

Therefore, $(1+ap)^{p^{n-1}} \equiv 1+ap^n \pmod{p^{n+1}}$. Thus, by induction,

$$(1+ap)^{p^{\alpha-2}} \equiv 1+ap^{\alpha-1} \pmod{p^\alpha} \text{ for all } \alpha \geq 2. \quad \square$$

Lemma 7.11: If $p \neq 2$ and $p \nmid a$, then the order of $(1+ap)$ in $\mathbb{Z}_{p^\alpha}^*$ is $p^{\alpha-1}$.

Proof: By Lemma 7.10, $(1+ap)^{p^{\alpha-1}} \equiv 1+ap^\alpha \pmod{p^{\alpha+1}}$. So, $(1+ap)^{p^{\alpha-1}} \equiv 1 \pmod{p^\alpha}$.

This implies that the order of $1+ap$ divides $p^{\alpha-1}$. On the other hand, by Lemma 7.9,

$$(1+ap)^{p^{\alpha-2}} \equiv 1+ap^{\alpha-1} \pmod{p^\alpha}. \text{ So, the order of } 1+ap \text{ does not divide } p^{\alpha-2} \text{ because}$$

$p \nmid a$. Therefore, the order of $1+ap$ is $p^{\alpha-1}$. $\square$

Theorem 7.12: Let p be an odd prime and α be a positive integer. Then $\mathbb{Z}_{p^\alpha}^*$ is cyclic.

Proof: By Theorem 7.5, $\mathbb{Z}_p^*$ is cyclic. If $g \in \mathbb{Z}_p^*$ such that $\langle g \rangle = \mathbb{Z}_p^*$, then $g + p$ also

generates $\mathbb{Z}_p^*$. If $g^{p-1} \equiv 1 \pmod{p^2}$, then $(g + p)^{p-1} \equiv g^{p-1} + p(p-1)g^{p-2} \pmod{p^2}$.

Since p^2 does not divide $p(p-1)g^{p-2}$, by replacing g with $g + p$ if necessary, we can assume that $g^{p-1} \not\equiv 1 \pmod{p^2}$.

To show that g generates $\mathbb{Z}_{p^\alpha}^*$, it suffices to show that if $g^n \equiv 1 \pmod{p^\alpha}$, then

$\phi(p^\alpha) = p^{\alpha-1}(p-1)$ divides n . So, suppose that $g^n \equiv 1 \pmod{p^\alpha}$. Notice that since

$g^{p-1} \equiv 1 \pmod{p}$ and $g^{p-1} \not\equiv 1 \pmod{p^2}$, $g^{p-1} = 1 + ap$ for some integer a , where $p \nmid a$.

Since $(1 + ap)^n \equiv 1 \pmod{p^\alpha}$, we have $p^{\alpha-1} \mid n$. Let $n = p^{\alpha-1} \tilde{n}$. Then

$g^n = (g^{p^{\alpha-1}})^{\tilde{n}} \equiv g^{\tilde{n}} \pmod{p}$. Thus, $g^{\tilde{n}} \equiv 1 \pmod{p}$. Since g generates $\mathbb{Z}_p^*$, $p-1$ divides

$\tilde{n}$. Therefore, $p^{\alpha-1}(p-1)$ divides n and $\mathbb{Z}_{p^\alpha}^*$ is cyclic. $\square$

The subsequent theorems are used to show, in Section 8, that the problem of factoring a composite integer N can be reduced to an order-finding problem that gives rise to an algorithm with a high probability of yielding a nontrivial factor. Theorems 7.13 and 7.21 are adapted and corrected from Nielsen and Chuang [NiCh]. Theorem 7.21 is proved using a version of the Chinese Remainder Theorem, which will be stated without proof.

Theorem 7.13: Let N be a composite integer and x be a solution to the equation

$x^2 \equiv 1 \pmod{N}$ such that $1 \leq x \leq N$ and $x \not\equiv \pm 1 \pmod{N}$. Then at least one of

$\gcd(x-1, N)$ and $\gcd(x+1, N)$ is a non-trivial factor of N .

Proof: By definition, $x^2 \equiv 1 \pmod{N}$ means that N divides $x^2 - 1 = (x+1)(x-1)$. Hence,

there exists an integer $d > 1$ such that d divides N and d divides $(x+1)$ or $(x-1)$. From

the assumption that $1 \leq x \leq N$ and $x \not\equiv \pm 1 \pmod{N}$, it follows that $x-1 < x+1 < N$, so

$d \neq N$. Therefore, at least one of $\gcd(x-1, N)$ and $\gcd(x+1, N)$ is a non-trivial factor

of N . $\square$

The next two lemmas are results from abstract algebra regarding the order of elements in a group. The proofs for Lemma 7.14 and 7.15 can be found in Fraleigh [Fral].

Lemma 7.14: Let G be a group. If the order of $x \in G$ is k , then the order of x^a is $k/\gcd(a, k)$.

Lemma 7.15: Let $G_1, G_2, \dots, G_m$ be groups. If $(g_1, g_2, \dots, g_m) \in G_1 \times G_2 \times \dots \times G_m$, then the order of $(g_1, g_2, \dots, g_m)$ is the least common multiple of the orders of the $g_i \in G_i$ for $i = 1, 2, \dots, m$.

Lemma 7.16: $\gcd(ab, ac) = a \gcd(b, c)$.

Proof: Let d be the greatest common divisor of ab and ac . Since a is a factor of both ab and ac , a divides d . So, $d = ak$ for some integer k . This implies that k must divide b and c . Moreover, k must be the greatest common divisor of b and c . Therefore,

$$d = \gcd(ab, ac) = ak = a \gcd(b, c). \quad \square$$

Theorem 7.17: If $C_k = \langle g \rangle$ is a cyclic group of order k , then for any $d|k$, $\varphi(d)$ is the number of elements of order d in C_k .

Proof: If $d|k$, $\gcd(a, d) = 1$, $0 < a < d$, then the order of $g^{a(k/d)}$ is d since

$$\gcd(ak/d, k) = (k/d)\gcd(a, d) = k/d. \text{ This gives at least } \varphi(d) \text{ elements of order } d$$

and accounts for $\sum_{d|k} \varphi(d) = k$ elements of C_k , i.e., all of them. Therefore, there are

exactly $\varphi(d)$ elements of order d . $\square$

Corollary 7.18: If C_k is cyclic of even order, then there exists a unique $x \in C_k$ such that the order of x is 2.

Proof: By Theorem 7.18, since 2 divides the order of C_k , there is $\varphi(2) = 2(1 - 1/2) = 1$

element of order 2 in C_k . $\square$

Corollary 7.19: Let p be an odd prime and $x \in \mathbb{Z}_p^*$. Then $x \equiv -1 \pmod{p^a}$ if and only if the order of x is 2.

Proof: Assume that $x \equiv -1 \pmod{p^a}$. Then $x \cdot x \equiv (-1) \cdot (-1) \equiv 1 \pmod{p^a}$, so the order of x is 2. By Corollary 7.18, this is the only element of order 2. $\square$

Corollary 7.20: Let C_k be cyclic of order $k = 2^m t$ for some odd integer t , and let l be such that $0 \leq l \leq m$. Then the number of elements of C_k of order $2^l s$, for some odd integer s , is less than or equal to $k/2$.

Proof: Fix l such that $0 \leq l \leq m$. The number of elements of order $2^l s$ for some odd integer s , is equal to

$$\begin{aligned} \sum_{s \nmid t} \phi(2^l s) &= \sum_{s \nmid t} \phi(2^l) \phi(s) \\ &= \sum_{s \nmid t} 2^{l-1} (1 - 1/2) \phi(s) \\ &= \sum_{s \nmid t} 2^{l-1} \phi(s) \\ &= 2^{l-1} \sum_{s \nmid t} \phi(s) \\ &= 2^{l-1} t. \end{aligned}$$

Given that $0 \leq l \leq m$, we have $2^{l-1} t \leq 2^{m-1} t = (2^m t)/2 = k/2$. $\square$

Chinese Remainder Theorem : Let n be a positive integer and $p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$ be its factorization into distinct primes. Then, as rings, $\mathbb{Z}_n \cong \mathbb{Z}_{p_1^{a_1}} \times \mathbb{Z}_{p_2^{a_2}} \times \dots \times \mathbb{Z}_{p_k^{a_k}}$. In particular, as multiplicative groups, $\mathbb{Z}_n^* \cong \mathbb{Z}_{p_1^{a_1}}^* \times \mathbb{Z}_{p_2^{a_2}}^* \times \dots \times \mathbb{Z}_{p_k^{a_k}}^*$.

From the Chinese Remainder Theorem, we see that there exists a bijective mapping from $\mathbb{Z}_n^*$ to $\mathbb{Z}_{p_1^{a_1}}^* \times \mathbb{Z}_{p_2^{a_2}}^* \times \dots \times \mathbb{Z}_{p_k^{a_k}}^*$ that sends an element x to $(x_1, x_2, \dots, x_k)$, where $x_i \equiv x \pmod{p_i^{a_i}}$.

We define an element of $x \in \mathbb{Z}_n^*$ of order r to be “bad” if r is odd or r is even and $x^{r/2} \equiv -1 \pmod{n}$. We say an element $x \in \mathbb{Z}_n^*$ is “good” if it is not bad.

Theorem 7.21: Let $n > 1$ be an odd integer with $n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$ for some integers a_i and distinct primes p_i for $1 \leq i \leq k$. Let $x \in \mathbb{Z}_n^*$ be such that $x \mapsto (x_1, x_2, \dots, x_k)$ and the order of x_i is $2^{e_i} s_i$, where s_i is an odd integer. Suppose the order of $x \in \mathbb{Z}_n^*$ is r .

Then the following are equivalent:

- (i) Either r is odd, or r is even and $x^{r/2} \equiv -1 \pmod{n}$.
- (ii) All of the e_i are equal.

Proof: By Lemma 7.16, the order of x is the least common multiple of the order of the x_i . Therefore, r is odd if and only if the order of each x_i is odd. This only occurs if each e_i is equal to 0. On the other hand, if r is even, then $r = 2^e s$, where $e = \max(e_1, e_2, \dots, e_k)$ and $s = \text{lcm}(s_1, s_2, \dots, s_k)$. Now, $r/2 = 2^{e-1} s$, so by Lemma 7.15,

$$\text{order of } x_i^{r/2} = \left(2^{e_i} / \gcd(2^{e_i}, 2^{e-1}) \right) (s_i / \gcd(s_i, s)) = \begin{cases} 2 & \text{if } e_i = e \\ 1 & \text{if } e_i < e. \end{cases} \quad (7.22)$$

Now, $x^{r/2} \equiv -1 \pmod{n}$ if and only if $x_i^{r/2} \equiv -1 \pmod{p_i^{a_i}}$ for all i . By Corollary 7.19, this occurs if and only if the order of $x_i^{r/2}$ is equal to 2 for all i . However, by equation 7.22, the order of $x_i^{r/2}$ is equal to 2 for all i if and only if $e_i = e$ for all i . $\square$

Corollary 7.23: Let N be a positive odd integer and $p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$ be its factorization into distinct primes. Then at most $\frac{1}{2^{k-1}}$ elements of $\mathbb{Z}_N^*$ are bad.

Proof: It is enough to show that for any $x_1 \in \mathbb{Z}_{p_1^{a_1}}^*$, at most $\frac{1}{2^{k-1}}$ of all $(x_2, x_3, \dots, x_k)$ give rise to a bad element $(x_1, x_2, \dots, x_k)$. If the order of $x_1 = 2^l s$ (s odd), then for $(x_1, x_2, \dots, x_k)$ to be bad, we must have $\text{order}(x_i) = 2^l t$, t odd, by Theorem 7.21.

However, by Corollary 7.20, at most $1/2$ of all $x_i \in \mathbb{Z}_{p_i^{a_i}}$ have order $2^l t$ (t odd).

Therefore, at most $\frac{1}{2^{k-1}}$ of all $(x_2, x_3, \dots, x_k)$ have all x_i with order $2^l t$. $\square$

8 Quantum Factoring Algorithm

In 1994, Peter Shor demonstrated that the problem of finding prime factors of a given integer n could be solved efficiently on a quantum computer. Shor's algorithm takes $O((\log n)^2 (\log \log n) (\log \log \log n))$ steps on a quantum computer and a polynomial in $\log n$ amount of post-processing time on a classical computer [Shor]. Currently, the fastest known general purpose classical factoring algorithm is the number field sieve derived from the works of T. El Gamal and proposed by J.M. Pollard in 1988, which requires $\exp(\Theta(n^{1/3} \log^{2/3} n))$ operations. As a consequence of Shor's algorithm, there are quantum algorithms which can factor a number exponentially faster than the best known classical algorithms.

The quantum factoring approach given in this paper can be found in [NiCh] and [Shor]. This quantum factoring algorithm utilizes the Euclidean algorithm for finding the greatest common divisor of two numbers and the continued fractions algorithm for expressing real numbers in terms of integers. We will reference these algorithms without descriptions. Detailed explanations of the Euclidean algorithm and the continued fractions algorithm can be found in [Ros].

Let N be a positive L -bit integer. We begin by reducing the problem of finding a non-trivial factor of N to an order-finding problem. If N is even, then 2 is a factor of N and we are done. Henceforth, we will assume that N is odd. The next step is to determine whether $N = y^a$ for some integers $y \geq 1$ and $a \geq 2$. This can be done in $O(L^3)$ time using a classical algorithm [NiCh]. If $N = y^a$, then y is a factor of N . For

$N \neq y^a$, we randomly choose an x such that $1 \leq x \leq N-1$. Then we compute the greatest common divisor of x and N using the Euclidean algorithm in $O(L^3)$ classical steps [NiCh]. If $\gcd(x, N) > 1$, then $\gcd(x, N)$ is a factor of N , and $k \geq 2$, since $N \neq y^a$. Otherwise, let r be the order of x in $\mathbb{Z}_N^*$. Suppose we can find r . If r is even and $x^{r/2} \not\equiv -1 \pmod{N}$, then we compute $\gcd(x^{r/2} - 1, N)$ or $\gcd(x^{r/2} + 1, N)$ and test to see if one of these is a non-trivial factor of N . By Corollary 7.23, the probability that both of these factors are trivial is at most $(\frac{1}{2})^{k-1}$ where k is the number of distinct prime factors of N . Therefore, after m iterations, the probability of obtaining a non-trivial factor is at least $1 - (\frac{1}{2})^m$, which can be made as close to 1 as desired.

We will now give a description of the order-finding algorithm. A depiction of the order-finding circuit is given in Figure 8.2 on page 67. The purpose of the order finding algorithm is to determine the order of x in $\mathbb{Z}_N$, i.e., to find the least positive integer r such that $x^r \equiv 1 \pmod{N}$. Recall that the number of elements in $\mathbb{Z}_N$ that are relatively prime to N is denoted by $\phi(N)$. It is obvious that $\phi(N) < N$. By Euler's Theorem [IrRo], since x is relatively prime to N , $x^{\phi(N)} \equiv 1 \pmod{N}$. Therefore, $r \leq N$.

We start by letting $t = 2L + 1$ and $q = 2^t$. Notice that $N^2 \leq q = 2^t$. The order-finding algorithm requires two registers. The first register contains t qubits initialized to $|0\rangle$. The second register contains L qubits initialized to $|0\rangle$.

Next, we put the first register in a superposition of states. This is accomplished by applying the Hadamard gate to each qubit in the first register. Therefore, we have $t = O(L)$ Hadamard gates. Once the gates are applied, the state of the machine is

$$\frac{1}{q^{t/2}} \sum_{a=0}^{q^t-1} |a\rangle |0\rangle.$$

Now, we need to give a method for constructing a classical reversible gate array to perform modular exponentiation. That is, we need to compute $x^a \pmod{N}$ from L -bit numbers a , x , and N . This is done by first computing $x^{2^i} \pmod{N}$ for all $i < L$ by repeated squaring. Since we need to compute $x^{2^i} \pmod{N}$ for all i up to $t-1$, this requires a total of $t-1 = O(L)$ squaring operations. Each squaring requires $O(L^2)$ steps for a total of $O(L^3)$ steps in the first stage. In the second stage, we multiply the powers $x^{2^i} \pmod{N}$ where 2^i appears in the binary expansion of a together to obtain $x^a \pmod{N}$. We need to perform at most $t-1$ modular multiplications, each requiring $O(L^2)$ steps, so the product can be computed in $O(L^3)$ steps. Hence, a reversible gate array that computes $x^a \pmod{N}$ from L -bit numbers a , x , and N would require $O(L^3)$ classical gates.

We quantize this method (Theorem 5.5) to obtain a quantum circuit that performs

$$|a\rangle |0\rangle \rightarrow |a\rangle |x^a \pmod{N}\rangle$$

in the second register. This leaves our machine in the state $\frac{1}{q^{1/2}} \sum_{a=0}^{q-1} |a\rangle |x^a \pmod N\rangle$.

Then we perform the quantum Fourier transform on the first register. Recall that the quantum Fourier transform sends $|a\rangle$ to $\frac{1}{q^{1/2}} \sum_{c=0}^{q-1} E(ac/q) |c\rangle$. Hence, our machine is now in the state

$$\frac{1}{q} \sum_{a=0}^{q-1} \sum_{c=0}^{q-1} E(ac/q) |c\rangle |x^a \pmod N\rangle. \quad (8.1)$$

Finally, we measure both registers in the basis $|x\rangle \otimes |y\rangle$, where $0 \leq x, y \leq q-1$, i.e., observe the value of $|c\rangle$ in the first register and $|x^a \pmod N\rangle$ in the second register.

Figure 8.2 [NiCh]. Quantum circuit for the order-finding algorithm.

Now, the probability that the order-finding algorithm ends in a state

$|c, x^k \pmod N\rangle$, $0 \leq k < r$, is given by taking $|z|^2$, where z is the coefficient of

$|c, x^k \pmod N\rangle$. The coefficient z is obtained by summing over all ways to reach the state $|c, x^k \pmod N\rangle$. Thus from equation 8.1, this probability is

$$|z|^2 = \left| \frac{1}{q} \sum_{a: x^a = x^k} E(ac/q) \right|^2, \quad (8.3)$$

where the sum is over all a , $0 \leq a < q$, such that $x^a \equiv x^b \pmod{N}$. Assuming that the order of x is r , this is the same as summing over all a satisfying $a \equiv k \pmod{r}$. If we rewrite a as $br + k$, expression 8.3 becomes

$$\begin{aligned} \left| \frac{1}{q} \sum_{b=0}^{\lfloor \frac{q-k-1}{r} \rfloor} E\left(\frac{(br+k)c}{q}\right) \right|^2 &= \left| \frac{1}{q} \sum_{b=0}^{\lfloor \frac{q-k-1}{r} \rfloor} E\left(\frac{brc}{q}\right) E\left(\frac{kc}{q}\right) \right|^2 \\ &= \left| \frac{1}{q} E\left(\frac{kc}{q}\right) \sum_{b=0}^{\lfloor \frac{q-k-1}{r} \rfloor} E\left(\frac{brc}{q}\right) \right|^2 \\ &= \left| \frac{1}{q} \sum_{b=0}^{\lfloor \frac{q-k-1}{r} \rfloor} E\left(\frac{brc}{q}\right) \right|^2 \end{aligned}$$

since $E\left(\frac{kc}{q}\right)$ has magnitude 1.

Let $\overline{rc}$ be the residue class of $rc \pmod{q}$, in the range $(-q/2, q/2]$. By replacing rc with $\overline{rc}$, we get

$$\left| \frac{1}{q} \sum_{b=0}^{\lfloor \frac{q-k-1}{r} \rfloor} E\left(\frac{b\overline{rc}}{q}\right) \right|^2.$$

Following the argument in [Shor], we can turn this sum into the following integral:

$$\frac{1}{q} \int_0^{\lfloor \frac{q-k-1}{r} \rfloor} E(b\overline{rc}/q) db + O\left(\frac{(q-k-1)/r}{q} E(\overline{rc}/q) - 1\right). \quad (8.4)$$

When $|\overline{rc}| \leq q/2$, the error term in expression 8.4 is bounded by $O(1/q)$ [Shor]. To

estimate the integral, we substitute u for rb/q . So, expression 8.4 becomes

$$\frac{1}{r} \int_0^{\lfloor \frac{r+1}{r} \rfloor} E\left(\frac{rc}{r}u\right) du.$$

If $(\overline{rc})/r$ varies between $-\frac{1}{2}$ and $\frac{1}{2}$, the absolute magnitude of the above integral is minimized when $\frac{\overline{rc}}{r} = \pm\frac{1}{2}$. Specifically, if $\frac{\overline{rc}}{r} = \pm\frac{1}{2}$, the integral approaches $2/(\pi r)$ as N , and therefore q , approach ∞ . The square of $2/(\pi r)$ is therefore an asymptotic lower bound on the probability that the final output is any particular state $|c, x^t \pmod{N}\rangle$ with $\overline{rc} \leq r/2$. So, the probability is asymptotically bounded below by $4/(\pi^2 r^2)$. Notice that $4/(\pi^2 r^2) > 1/(3r^2)$, so the probability is at least $1/(3r^2)$ for sufficiently large N [Shor].

Summarizing, we see that the probability that the final state is $|c, x^t \pmod{N}\rangle$ will be at least $1/(3r^2)$ if

$$-\frac{r}{2} \leq \overline{rc} \leq \frac{r}{2},$$

i.e., if there exists an s such that

$$\begin{aligned} -\frac{r}{2} &\leq rc - sq \leq \frac{r}{2} \\ \left(-\frac{r}{2} \div rq\right) &\leq \frac{rc - sq}{rq} \leq \left(\frac{r}{2} \div rq\right) \\ -\frac{1}{2q} &\leq \frac{c}{q} - \frac{s}{r} \leq \frac{1}{2q} \\ \left|\frac{c}{q} - \frac{s}{r}\right| &\leq \frac{1}{2q} = 2^{-t-1}, \end{aligned}$$

i.e., the probability of the final state $|c, x^t \pmod{N}\rangle$ is at least $1/(3r^2)$ if c/q is a t -bit approximation to s/r .

We now see that this “good” result $\left| (c/q) - (s/r) \right| \leq 1/(2q)$ occurs for each state $|c, x^k \pmod{N}\rangle$ such that s/r is close to c/q with probability greater than or equal to $1/(3r^2)$. Moreover, there is one such state for each $x^k \pmod{N}$ and each c/q near s/r for some s such that $0 \leq s \leq r$. Hence, the total probability of a “good” result is at least

$$\frac{1}{3r^2} (r \text{ values of } s) (r \text{ values of } x^k \pmod{N}) = \frac{1}{3}.$$

So, the probability of obtaining a “bad” result after k iterations is less than or equal to $(\frac{2}{3})^k$, which can be made arbitrarily close to 0 with a constant number, k , of iterations.

If we get a “good” result, the following theorem may then be used to compute s/r from c/q .

Theorem 8.5 [NiCh]: Suppose s/r is a rational number such that $\left| \frac{s}{r} - \mathcal{G} \right| \leq \frac{1}{2r^2}$. Then

s/r is a convergent of the continued fraction expansion of $\mathcal{G}$, and thus can be computed in $O(L^3)$ classical operations using the continued fractions algorithm.

We already know c and q . Let $\mathcal{G} = c/q$, then by Theorem 8.3, s/r is a convergent of $\mathcal{G}$ since $q > N^2 \geq r^2$ and $\frac{1}{2q} \leq \frac{1}{2r^2}$. Since $q > N^2$, there is a unique fraction that satisfies the inequality $\left| (s/r) - (c/q) \right| \leq 1/(2q)$ with $r < N$. So, by rounding c/q to the nearest fraction with denominator less than N , we can obtain s/r in lowest terms by using the continued fractions algorithm.

At this point, we only get s/r in reduced terms, so we are only guaranteed to get a denominator that divides r , not r itself. So then, we can verify that the candidate for r is the order of x by checking to see if $x^r \equiv 1 \pmod{N}$. (Note that we will not obtain a false r since the candidate is less than or equal to the order of x .) If r is not the order of x , then we repeat the order-finding algorithm.

So how often do we get s/r with reduced terms s and r , i.e. how often is $\gcd(s, r) = 1$? Well, we know that there are r possible values for x^t . Hence, there are $r\phi(r)$ states $|c, x^t \pmod{N}\rangle$ from which we may obtain r (one for each s value and x^t value). Since each of these states occurs with probability less than or equal to $1/(3r^2)$, we can obtain r with probability at least $\phi(r)/(3r)$. Furthermore, it can be shown that $\phi(r)/r > \varepsilon/(\log \log r)$ for some positive constant ε [HaWr]. Thus, the probability of obtaining the correct r is at least $\varepsilon/(\log \log r)$. This means that we have a very high probability of success by repeating the order-finding algorithm $O(\log \log r)$ times [Shor].

Theorem 8.6: Let N be a positive integer and L be the number of bits needed to represent N . The order-finding algorithm described in this section requires a total of $O(L^3 \log L)$ gates to obtain the order of an element $x \in \mathbb{Z}_N^*$ with a fixed probability $1 - \varepsilon$.

Proof: The Hadamard transform performed on the first register requires $O(L)$ quantum gates. Computing $x^a \pmod{N}$ by modular exponentiation uses $O(L^3)$ quantum gates.

By Theorem 6.8, the quantum Fourier transform can be done with $O(L^2)$ quantum gates.

The continued fractions algorithm adds another $O(L^3)$ classical gates. Since we only need repeat this procedure $O(\log \log r) \leq O(\log \log N) = O(\log L)$ times, to get the fixed desired probability, the total cost for implementing the order-finding algorithm is $O(L^3 \log L)$. $\square$

Corollary 8.7: The factoring algorithm described in this section uses $O(L^3 \log L)$ quantum and classical gates.

Proof: We can determine whether N is even or $N = y^a$ for some integers $p \geq 1$ and $a \geq 2$ using $O(1)$ and $O(L^3)$ operations, respectively. The third step utilizes the Euclidean algorithm to return a factor of N or a randomly chosen element x , $0 \leq x \leq N-1$. The fourth step uses the order-finding algorithm. The fifth step again utilizes the Euclidean algorithm to find a non-trivial factor of N . Now, the Euclidean algorithm can be performed using $O(L^3)$ operations, and by Theorem 8.6, the order-finding algorithm uses $O(L^3 \log L)$. Therefore, the factoring algorithm uses a total of $O(L^3 \log L)$ quantum and classical gates. $\square$

Currently, the most widely used public key cryptosystem is the RSA cryptosystem developed by Ronald Rivest, Adi Shamir, and Leonard Adleman [Ros]. For RSA to be secure, it must be impossible to factor large numbers in a short period of time. However, since Shor's algorithm is exponentially faster than any known classical method, a

quantum computer could be used to break RSA. This fact has stimulated much research in the field of quantum information and computation.

References

- [FIS] S.H. Friedberg, A.J. Insel, and L.E. Spence, *Linear Algebra*. Prentice Hall Inc., 1997.
- [Fral] J.B. Fraleigh, *Abstract Algebra*. Addison-Wesley Publishing Company, Inc., 1999.
- [Fran] A. Franklin, *Stanford Encyclopedia of Philosophy Supplement to Experiment in Physics* [book on-line], “Appendix 5: Right Experiment, Wrong Theory: The Stern Gerlach Experiment.” Available at <http://plato.stanford.edu/entries/physics-experiment/app5.html>; Internet; accessed 20 February 2002.
- [HaWr] G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*. Oxford University Press, 1979.
- [HoJo] R.A. Horn and C.R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [Hung] T.W. Hungerford, *Algebra*. Springer-Verlag New York Inc., 1974.
- [IrRo] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*. Springer-Verlag New York Inc., 1982.

- [Meg] Z. Meglicki, *Introduction to Quantum Computing (M743)* [course notes on-line]. Available from <http://beige.ucs.indiana.edu/B679/>; Internet; accessed 6 February 2002.
- [NiCh] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [Rae] A.I.M. Rae, *Quantum Mechanics*. McGraw-Hill Book Company (UK) Limited, 1981.
- [Ros] K.H. Rosen, *Elementary Number Theory and Its Applications*. Addison-Wesley Publishing, 2000.
- [Shor] P.W. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer." *SIAM Journal of Computing*, 26 (1997): 1484-1509.