Quantum Mechanics, Quantum Computing and Heat Computers: An Introduction

J. D. Vélez-J.P. Hernandez-D.J. Gómez

July 31, 2023

Preface

In these notes, we provide a rather idiosyncratic introduction to both Quantum Mechanics and Quantum Computing. Most book, with a few exceptions (see, for instance, [C-T] and [Gr]) discuss each subject separately. This becomes an obstacle for many mathematicians who want to learn Quantum Computing, not only as an axiomatic theory, but as a theory in which the physical principles are also explained and understood, so that it becomes possible to understands the many implementations of quantum computers in the real world. This notes also include new ideas and new research that has been carried out recently by one of the authors ([TCM]) in which a new paradigm of computation, in many ways similar to quantum computing, is explored. A complete description of the famous Shor's algorithm is presented from this new view point.

The course is divided in six chapters. The first chapter is an introduction to the basic ideas of Quantum Mechanics. In the second chapter, we introduce finite dimensional systems of particles and discuss the famous EPR experiment and the phenomenon of entangled particles. In the third chapter, we introduce the formalism of quantum theory and present its mayor applications to finite dimensional systems.

In the fourth chapter, we give a brief introduction to quantum computing. As the main example, we discuss Shor's algorithm.

Chapter five deals with infinite dimensional systems. We discuss what was the original motivation for the formulation of Quantum Mechanics: Schrödinger's famous wave equation. After introducing the Wave Equation, we present some of its main application.

In the final chapter, we introduce the concept of *random computing* as a similar alternative to quantum Computing. As a demonstration of its power, we prove that an analogue of Shor's algorithm that also runs in polynomial time can also be implemented within this new paradigm of computation.

We assume no previous knowledge from the reader. These notes should be regarded as a first draft of a future book on these subjects, and are still in an editorial process. They do not pretend to be a complete account of such vast subjects as Quantum Mechanics or Quantum Computing, but rather as a short introduction in which we have tried to motivate as much as possible the main ideas as well as the mathematical formalism lying behind these extraordinary theories.

Content

Chapter 1

Introduction to Quantum Mechanics

We start by analyzing the celebrated experiment of Young with double slits to validate the undulatory theory of light. Then we will see how the photoelectric effect suggests a corpuscular nature of light and the need for a new paradigm. Then we will study the polarization of light from a classical perspective. Explaining these experiments with the corpuscular theory will be our way to introduce the quantum model.

1.1 Maxwell's equations

The undulatory theory of light was established by Young as early as 1803 [Young]. Many scientists had proposed a wave theory of light based on experimental observations. Nevertheless, Isaac Newton had rejected the wave theory and adopted instead a corpuscular theory according to which light is emitted from a luminous body in the form of very tiny particles. This theory was accepted by most scientists, including Laplace and Biot despite the fact that many phenomena, like diffraction, could not be adequately explained by such a theory.

Young's famous experiment is the predecessor of the double slit experiment. In its modern form the experiment can be described as follows: a monochromatic source of light, like a laser, illuminates a plate of metal or some other material pierced by two parallel slits at distance d. The light passing through the slits is then projected on a screen behind the plate, located at distance L. Light waves passing through the two slits interfere producing bright and dark fringes on the screen, an outcome that would not be expected if light consisted of classical particles (see Figures 1.1[1] and 1.2[2])

This phenomenon is called *interference*. When two identical waves interact they may increase their amplitudes or sometimes cancel each other. Consider how two waves travel from their corresponding slits to the screen, as illustrated in Figure 1.3. The wave that starts at B travels a distance l_1 which is less than the distance l_2 traveled by the wave originating at A. Hence, different numbers of wavelengths fit into each path. The two waves start out from their slits in phase, but they may end up out of phase at the screen. This occurs if the paths differ in length by half a wavelength. In this case we say that they interfere *destructively*. If the paths differ by a whole wavelength, the waves arrive in phase, interfering *constructively*.

[1] Source: http://www.cobocards.com/pool/de/card/7puji0113/online-karteikarten-beugung-und-interferenz-am-doppelspalt/

[2] Source: https://courses.lumenlearning.com/physics/chapter/27-3-youngs-double-slit-experiment/

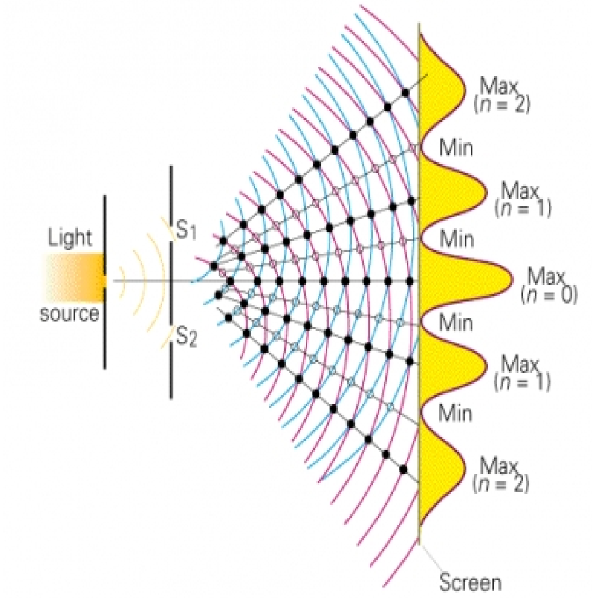


Figure 1.1: Young's experiment.

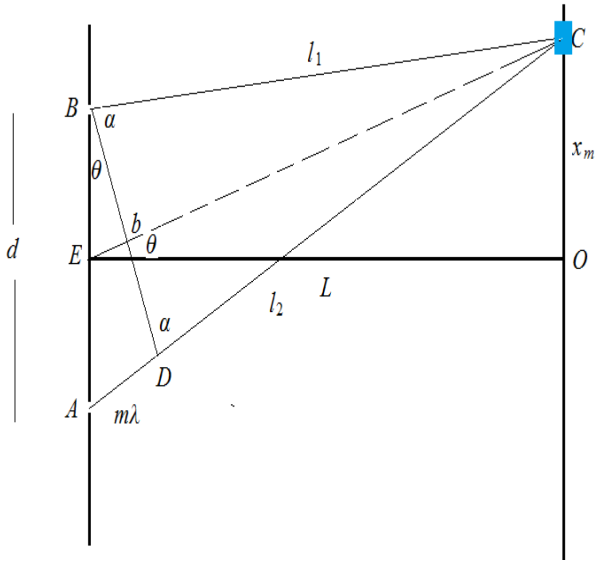


Figure 1.2

More generally, destructive interference occurs if the paths taken by the two waves differ by any half-integral number of wavelengths $(1/2)\lambda$, $(3/2)\lambda$, $(m+1/2)\lambda$... On the other hand, constructive interference occurs in case they differ by any integral number of wavelengths λ, $2\lambda, ..., m\lambda$ ...

In Figure 1.2 we can see that the angle α is very close to $\pi/2$. This is because d is small (typically less that a millimeter), and, for the first fringes (small m), $m\lambda$ is also very small (λ is of the order of nanometers). Therefore $b \leq m\lambda + d$ is also very small.

Then the isosceles triangle BDC has very large height compared to its base. This means that the opposite angle to side b (BCD) is almost zero. Thus, $2\alpha = \pi$ and therefore $\alpha = \pi/2$. As a consequence of this we may assume that BD is approximately perpendicular to AC and therefore the angle CEO is also very close to θ.

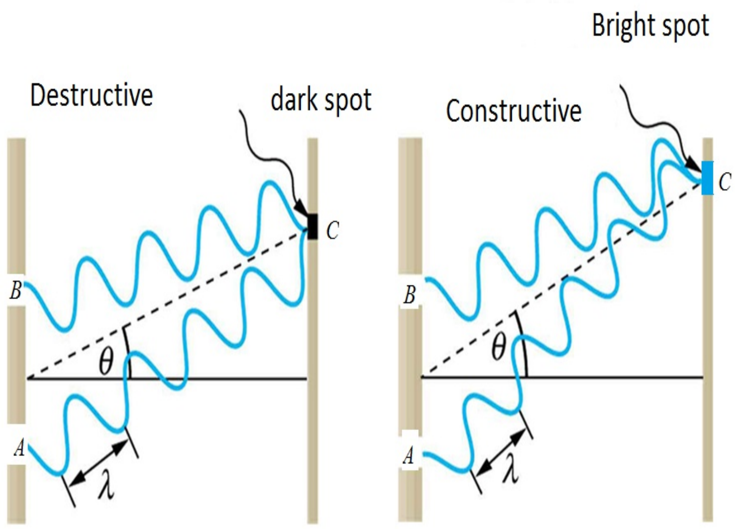


Figure 1.3

From those approximations one obtains an estimate for the difference $l_2 - l_1$, and for x_m.

$$l_2 - l_1 = m\lambda = d \sin\theta,$$

$$\frac{x_m}{L} = \tan\theta \approx \theta.$$

Henceforth,

$$x_m = \frac{m\lambda L}{d}.$$

For instance, let us consider the case of a green laser, with wavelength equal to $\lambda = 520 \times 10^{-9}$ m. Suppose the two slits in our experiment are separated by a distance $\delta = 0.2543 \times 10^{-3}$ m and suppose that $L = 2.14$ m is the distance from the slits to the screen. We see in this case that $x_1 = 4.37$ mm and $x_2 = 8.7$ mm.

1.2 The undulatory model of light

In classical optics light is an electromagnetic wave with a frequency in the visible spectrum. A typical human eye will respond to wavelengths from about 380 to about 750 nanometers which corresponds to a frequency in the band of $400{\times}10^{12}$–$790{\times}10^{12}$ Hertz. Light propagating in the vacuum is modeled by two time dependent vector fields $E(r,t)$ and $B(r,t)$, an electric and a magnetic field, where we denote by (r) coordinates for $\mathbb{R}^3$ and t denotes the time coordinate. Maxwell's equations in the vacuum take the following form:

$$\begin{aligned}
&\mathsf{div}E = 0\\
&\mathsf{div}B = 0\\
&\mathsf{RotE} = -\frac{\partial B}{\partial t}\\
&\mathsf{RotB} = \epsilon_0\mu_0\frac{\partial E}{\partial t}
\end{aligned}$$

where ϵ_0 and μ_0 are two constant called the *permitivity* and the *permeability* of the vacuum. From Maxwell's equations we derive the *wave equation* for $E(r,t)$ and $B(r,t)$. For this we use the *Laplacian identity.* For any vector field W one has

$$\mathsf{Rot}(\mathsf{Rot}W) = \mathsf{grad}(\mathsf{div}W) - \nabla^2 W.$$

where $\nabla^2 W$ is the Laplacian of W. Since $\mathsf{div}E = 0$ and $\mathsf{Rot}E = -\partial B/\partial t$ one obtains

$$\begin{aligned}
\mathsf{Rot}(\mathsf{Rot}E) &= \mathsf{Rot}(-\frac{\partial}{\partial t}B)\\
&= -\frac{\partial}{\partial t}\mathsf{Rot}(B) = -\nabla^2 E.
\end{aligned}$$

On the other hand,

$$\mathsf{RotB} = \epsilon_0\mu_0\frac{\partial E}{\partial t}.$$

Henceforth,

$$-\nabla^2 E + \epsilon_0\mu_0\frac{\partial^2 E}{\partial t^2} = 0. \tag{1.1}$$

Or in more standard form

$$\nabla^2 E_i = \frac{1}{c^2}\frac{\partial^2 E_i}{\partial t^2},$$

where E_i are the coordinates of E and

$$c = \frac{1}{\sqrt{\epsilon_0\mu_0}}.$$

Similarly one derives the corresponding equation for each coordinate B_i of the magnetic field

$$\nabla^2 B_i = \frac{1}{c^2}\frac{\partial^2 B_i}{\partial t^2}.$$

One particular solution of the wave equation corresponds to a monochromatic wave with frequency ω that travels along the z-axis. In this case a solution is given by

$$E(x,y,z,t) = A\cos(2\pi\omega t - \delta_1)e_1 + B\cos(2\pi\omega t - \delta_2)e_2, \tag{1.2}$$

where $\delta_1 = \frac{\omega}{c}z + a$ and $\delta_2 = \frac{\omega}{c}z + b$ for arbitrary constants a y and b. Here e_1, e_2, e_3 denote the standard vectors of $\mathbb{R}^3$ in the spatial directions of the Cartesian coordinates x, y, z.

This particular solution only depends on z and t. If we fix z we may choose a and b such that $\frac{\omega}{c}z + a = \frac{\omega}{c}z + b = 0$. This amounts to set an appropriate origin for the z and t coordinates where the equation above then looks like

$$\begin{aligned} E_z(t) &= A\cos(2\pi\omega t)e_1 + B\cos(2\pi\omega t)e_2, && (1.3)\\ &= (Ae_1 + Be_2)\cos(2\pi\omega t). && (1.4)\end{aligned}$$

The vector $n = Ae_1 + Be_2$ may be written as $n = \sqrt{I}\mathbf{n}_\theta$ where $\mathbf{n}_\theta = \cos(\theta)e_1 + \sin(\theta)e_2$ is a unitary vector in the direction of n_θ, $0 \leq \theta \leq \pi$ is the angle n determines with respect to the x-axis and $I = A^2 + B^2$. With this notation we may write

$$E_z(t) = \sqrt{I}\cos(2\pi\omega t)\ \mathbf{n}_\theta. \qquad (1.5)$$

The square of the norm $|E_z(t)|^2 = I\cos^2(2\pi\omega t)$ represents the *intensity* of the light wave at time t in the plane $z =$ constant. We see from (1.5) that the electric field (similarly for the magnetic filed) oscillates in a constant plane. In this case we say that the beam of light is *polarized linearly* in the direction of $\mathbf{n}_\theta$.

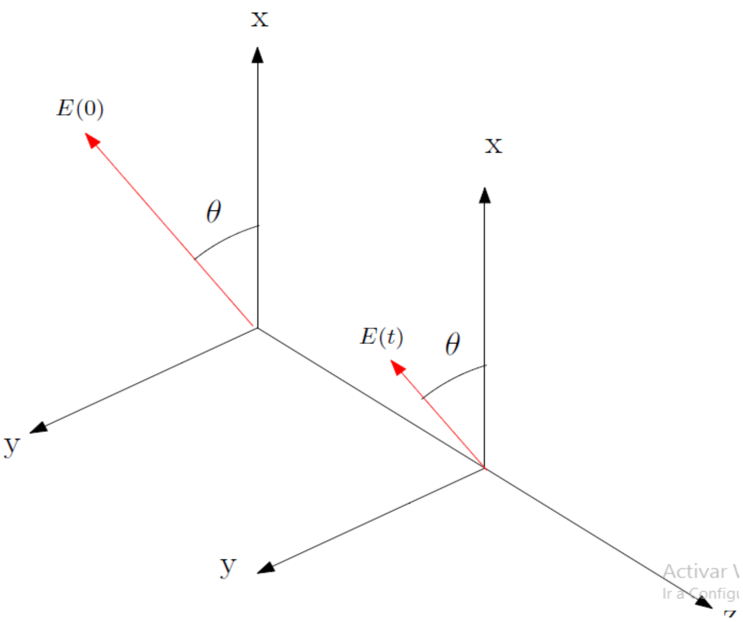


Figure 1.4: Electric Field

1.3 Corpuscular model

In the photoelectric effect electrons are emitted when electromagnetic radiation, such as light, impinges on a particular type of material. The first scientist who studied the phenomenon noticed that many experimental results disagreed with the predictions of classical electromagnetism. For instance, an alteration in the intensity of light would theoretically change the kinetic energy of the emitted electrons but the experiments showed instead that electrons were dislodged only when light exceeds a certain frequency, regardless of the light's intensity or time exposure.

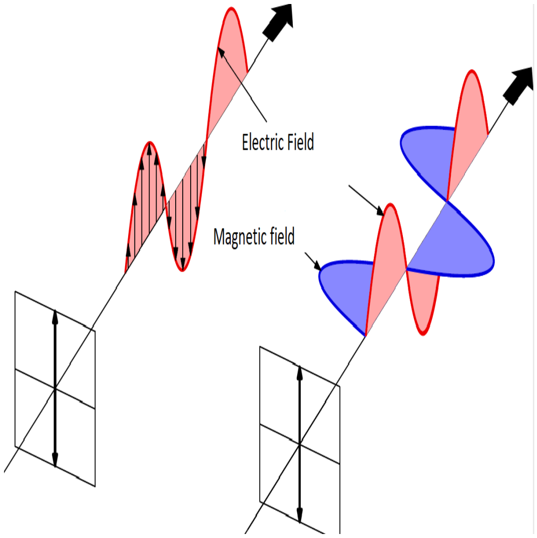


Figure 1.5: Electric and Magnetic Fields

In 1900, while studying black-body radiation, the German physicist Max Planck suggested that the energy carried by electromagnetic waves could only be released in packets of energy. In 1905, Albert Einstein proposed that a beam of light is not a wave propagating through space, but a swarm of discrete energy packets, known as *photons*. The amount of energy one can obtain from a monochromatic source of light comes in discrete units of energy. If we denote by ω the frequency of the light impinging on the material, the total energy is given by an integer multiple of $\hbar\omega$, where $\hbar$ (h bar) is equal to Planck's constant $h = 6,62 \times 10^{-34}$ J × s divided by 2π.

The discovery of the photoelectric effect was a key step in the development of quantum mechanics.

1.4 Polarization of light

As we observed above, one particular solution to the wave equation (1.1) corresponds to linearly polarized light. That is, light whose oscillating electric field points in the direction of a vector $\mathbf{n}_\theta$ (Equation 1.3).

Most common sources of visible light, such as the sun, flames, and incandescent lamps, consist of an equal mixture of polarizations. Polarized light, nonetheless, can be obtained by using certain crystals that show a property called *dichroism* or a preferential absorption of light which is polarized in some directions.

The most common method for obtaining linearly polarized light involves the use of a *Polaroid filter*. When unpolarized light is transmitted through a Polaroid filter, its electric field $E_z(t)$ is *projected* in the direction of $\mathbf{n}_\theta$ (see 1.5). Mathematically, we call this interpretation as the *projection principle*. Any polarizer capable of polarizing light linearly in the direction of $\mathbf{n}_\theta$ will be denoted by $P(\theta)$.

According to this principle, if we let light go sequentially trough two polarizes in a row, $P(\theta_1)$ and $P(\theta_1 + \pi/2)$, the first filter has the effect of projecting $E_z(t)$ over the unitary vector $\mathbf{n}_{\theta_1}$. Light that passes this filter will then be projected into an orthogonal direction, $\mathbf{n}_{\theta_1+\pi/2}$. Therefore no light will pass this second filter, as observed experimentally. More generally, if the second polarizer is of type $P(\theta_2)$, an empirical observation known as *Malus law* predicts that light will come out of $P(\theta_2)$ with intensity equal to $I_2 = I_1 \cos^2(\alpha)$, where I_1 is the intensity after passing $P(\theta_1)$, and $\alpha = \theta_2 - \theta_1$.

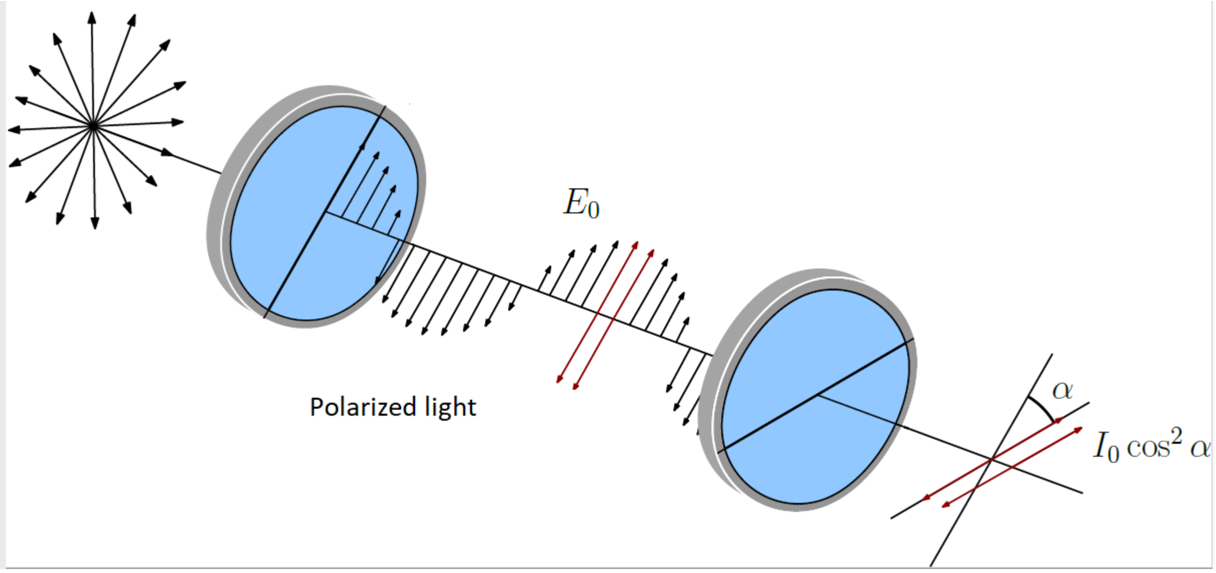


Figure 1.6: Polarization filters in a row

The undulatory model of light explains this phenomenon. In fact, after passing $P(\theta_1)$ the electric field will be (see 1.5)

$$E_z(t) = \sqrt{I}\cos(2\pi\omega t)\ \mathbf{n}_{\theta_1}.$$

Hence, $I_1 = I\cos^2(2\pi\omega t)$. The projection onto $\mathbf{n}_{\theta_2}$ is given by

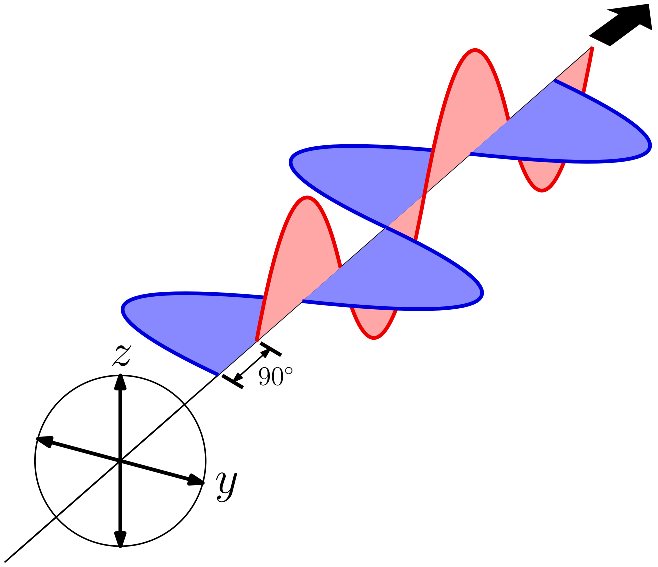
$$\begin{aligned}\langle E_z(t), \mathbf{n}_{\theta 2}\rangle\ \ \mathbf{n}_{\theta_2} &= \sqrt{I}\cos(2\pi\omega t)\,\langle \mathbf{n}_{\theta_1}, \mathbf{n}_{\theta_2}\rangle \\ &= \sqrt{I}\cos(2\pi\omega t)\cos(\alpha).\end{aligned}$$

We have shown that the intensity of light after passing $P(\theta_2)$ will be equal to $I_1\cos^2(\alpha)$, as predicted by Malus law (sse Figure 1.6).

1.5 Circularly polarized light

There is another solution to the wave equation known as circularly polarized light. In this case the electric and magnetic fields present a phase difference equal to $\pi/2$. At each point the electromagnetic field of the wave has a constant magnitude and is rotating at a constant rate in a plane perpendicular to the direction of the wave. Hence, at any instant, the electric field vector of the wave indicates a point on a helix oriented along the direction of propagation. A circularly polarized wave can rotate in one of two possible senses: clockwise or *right-handed circular polarization*, in which the electric field vector rotates in a right-hand sense with respect to the direction of propagation, and counter-clockwise or *left-handed circular polarization*, in which the vector rotates in a left-hand sense. These correspond to solutions of the wave equation where $\delta_1 = 0$ and $\delta_2 = \pm\pi/2$, respectively, and $A = B = \sqrt{I/2}$.

$$\begin{aligned} E_z(t) &= \frac{\sqrt{I}}{\sqrt{2}}\cos(2\pi\omega t)e_1 + \frac{\sqrt{I}}{\sqrt{2}}\cos(2\pi\omega t \pm \pi/2)e_2 \qquad (1.6)\\ &= \frac{\sqrt{I}}{\sqrt{2}}\cos(2\pi\omega t)e_1 \pm \frac{\sqrt{I}}{\sqrt{2}}\sin(2\pi\omega t)e_2. \end{aligned}$$



1.6 The corpuscular model and the polarization of light

The polarization of light can be well explained in the undulatory model. However, if we adopt the corpuscular model we immediately run into trouble. Firstly, the projection principle can no longer be applied, since a projection will change the total energy $E = \hbar\omega$ of a single photon by an arbitrary fraction. Hence, at any polarizing filter a photon must either be able to pass it or it is absorbed.

One would be tempted to consider a mathematical model in which each photon would internally carry a sequence of well determined "inner states of polarizations". One could model these states as a binary sequence of digits (λ_θ), $0 \leq \theta \leq 2\pi$. If $\lambda_\theta = 1$, this will be interpreted as the ability of that particular photon to go through a filter of type $P(\theta)$; and $\lambda_\theta = 0$, otherwise. For a particular photon we may ignore what this sequence is. However, the key point is that it is an already fixed property that only depends on the particular photon we are examining.

But this model cannot hold. It is observed that after light goes trough any $P(\theta)$ its intensity reduces to a half. This is well explained in the classical theory. In fact, it is reasonable to assume that monochromatic light of fixed intensity I comes in the form of a random uniform mixture of polarizations. That is, one would expect to find a mixture of different electric field of the form $E_\alpha = \sqrt{I}\mathbf{n}_\alpha$, for equally distributed values of $0 \leq \alpha \leq 2\pi$. Malus law would then predict that the average intensity of the light passing $P(\theta)$ would be given by

$$\frac{1}{2\pi}\int_0^{2\pi} I\cos^2(\alpha)d\alpha = \frac{I}{2\pi} \times \pi = \frac{I}{2}.$$

In our toy model this would mean that, on average, for each $0 \leq \alpha \leq 2\pi$, photons come in nature in equal proportions, half with $\lambda_\alpha = 1$ and the other half with $\lambda_\alpha = 0$.

But then let us try to see what would happen if monochromatic light of intensity I goes through a sequence of three polarizers in a row, $P(0)$, $P(\pi/4)$ and $P(0)$. According to the classical model, after passing the first filter light would be polarized in the direction of $\mathbf{n}_0 = e_1$ and its intensity would be $I/2$. After going through the second filter it will come out polarized in the direction of $\mathbf{n}_{\pi/4}$, with intensity $I/4$. After passing the third filter we will get light polarized in the direction of e_1, again, but with intensity equal to $I/8$.

In our model one half of the photons will pass the first polarizer. For these we know that $\lambda_0 = 1$. Then, only one quarter will go through the second filter. For those photons we know for sure that $\lambda_0 = \lambda_{\pi/4} = 1$. But then, when they encounter the third filter, all photons should pass, since they all have $\lambda_0 = 1$. Thus, the intensity of the beam of light after passing all three polarizers will be $I/4$, not $I/8$, contrary to the empirical evidence.

1.7 Quantum Model

The electric field of a linearly polarized monochromatic wave, $E_z(t) = \sqrt{I}\cos(2\pi\omega t)\mathbf{n}_\theta$, can be written as $E_z = \sqrt{I}Re(e^{-2\pi\omega t i})\mathbf{n}_\theta$. For circularly polarized light we may also write its electric field in complex form as $E_z(t) = \sqrt{I}Re(e^{-2\pi\omega i t})\mathbf{n}$, where

$$\mathbf{n} = \frac{1}{\sqrt{2}}e_1 \pm \frac{i}{\sqrt{2}}e_2. \tag{1.7}$$

In fact:

$$\sqrt{I}Re(e^{-2\pi\omega i t})\mathbf{n} = \frac{\sqrt{I}}{\sqrt{2}}\cos(2\pi\omega t)e_1 + \frac{\sqrt{I}}{\sqrt{2}}\sin(2\pi\omega t)e_2.$$

This suggests that to each linearly or circularly polarized photon we may associate a two dimensional complex vector that we may call its *polarization.* In general, for any photon that corresponds to a monochromatic wave whose electric field at each constant plane z is given by

$$E_z(t) = A\cos(2\pi\omega t - \delta_1)e_1 + B\cos(2\pi\omega t - \delta_2)e_2,$$

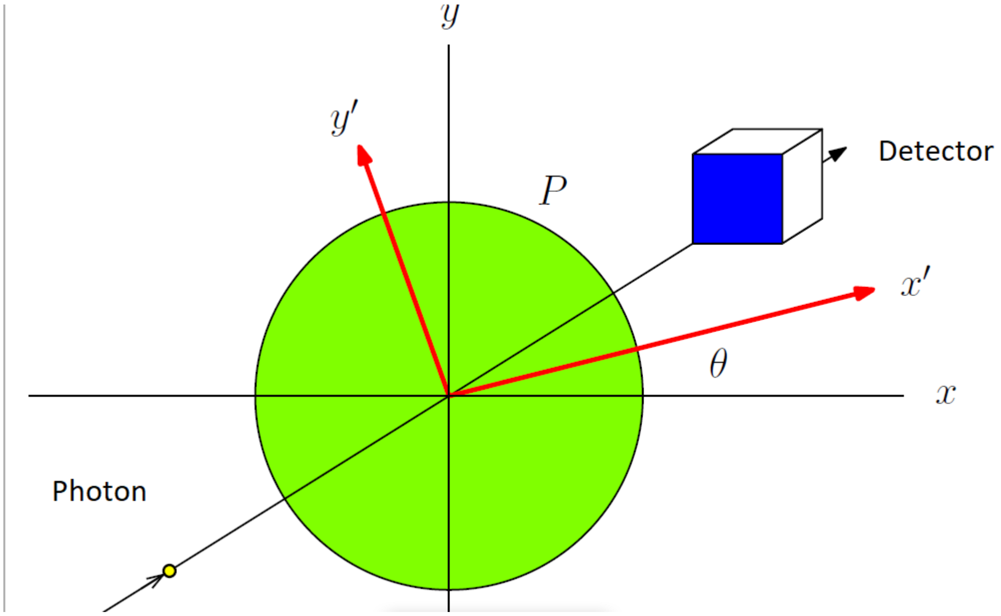
we define its *polarization* as the *unitary* vector (with norm equal to one) in $\mathbb{C}^2$ defined as:

$$\phi = \alpha\ e^{i\delta_1}e_1 + \beta\ e^{i\delta_2}e_2, \tag{1.8}$$

where $\alpha = A/(A^2+B^2)$, $\beta = B/(A^2+B^2)$. The coefficients $\alpha e^{i\delta_1}$ and $\beta e^{i\delta_2}$ are called the *complex amplitudes of* ϕ. The field $E_z(t)$ can be recovered as

$$E_z(t) = \sqrt{I}Re\left[e^{-2\pi\omega i t}(\alpha e^{i\delta_1}e_1 + \beta e^{i\delta_2}e_2)\right].$$

Let us analyze again what happens when a photon goes through a filter $P(\theta)$. We know it may happen that the photon is either absorbed or it passes the filter



Going through $P(\theta)$ has the effect of *collapsing* the polarization state into the vector $\mathbf{n}_\theta$. This collapse corresponds to a *measurement*, a process that entails two steps. Let us denote by $\beta_\theta = \{e_1(\theta), e_2(\theta)\}$ the orthonormal oriented base of $\mathbb{C}^2$ given by $e_1(\theta) = \mathbf{n}_\theta$ and by $e_2(\theta)$ its orthogonal complement.

1. Let $\phi = \alpha\, e^{i\delta_1} e_1 + \beta\, e^{i\delta_2} e_2$ be the polarization of the photon (1.8). Let $\phi = \alpha' e_1(\theta) + \beta' e_2(\theta)$, with $|\alpha'|^2 + |\beta'|^2 = 1$ be the expression of this vector in the basis β_θ. Then, *measuring the polarization at* $P(\theta)$ has the effect of collapsing ϕ into either $e_1(\theta)$, if the photon passes the filter or into $e_2(\theta)$ if it is absorbed. Hence, this measurement corresponds to a transition of states $\phi \longmapsto e_i(\theta)$.

2. The squared norms of the coefficients α' and β' are interpreted as *probabilities*: $|\alpha'|^2$ is the probability that the transition $\phi \longmapsto e_1(\theta)$ occurs while $|\beta'|^2$ is the probability of the transition $\phi \longmapsto e_2(\theta)$, the opposite event.

1.8 Filters in a row (revisited)

Let us test our model and see what happens when a photon encounters several filters in a row. Suppose these are $P(\theta_1)$ and $P(\theta_2)$. If we write its polarization state as $v = ae_1(\theta_1) + be_2(\theta_1)$, then the first *measurement* at $P(\theta_1)$ collapses v into two possible vectors: $v \longmapsto e_1(\theta_1)$ or $v \longmapsto e_2(\theta_1)$. On the other hand, let us write $e_i(\theta_1)$ in the orthonormal base $\{e_1(\theta_2), e_2(\theta_2)\}$

$$e_i(\theta_1) = d_i e_1(\theta_2) + c_i e_2(\theta_2).$$

Passing through the first filter amounts to a transition $v \longrightarrow e_1(\theta_1)$. If this occurs, the second measurement at $P(\theta_2)$ collapses the vector $e_1(\theta_1) = d_1 e_1(\theta_2) + c_1 e_2(\theta_2)$ into any of the two vectors $e_1(\theta_2)$ or $e_2(\theta_2)$ with probabilities $|d_1|^2$ and $|c_1|^2$, respectively.

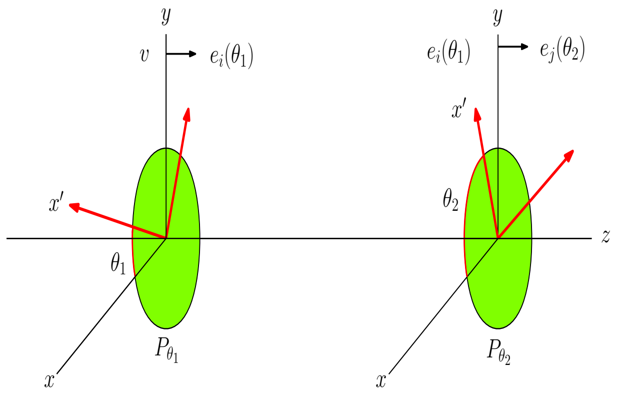


Figure 1.7: polarizers in a row

For instance, if $\theta_1 = 0$ and $\theta_2 = \pi/2$ then it is not possible for any photon to pass through both filters. After passing the first one we know that its polarization state would be $e_1 = e_1(0)$. On the other hand, $e_1(\pi/2) = e_2$ and $e_2(\pi/2) = -e_1$. Since

$$e_1(0) = 0e_1(\pi/2) + (-1)e_2(\pi/2),$$

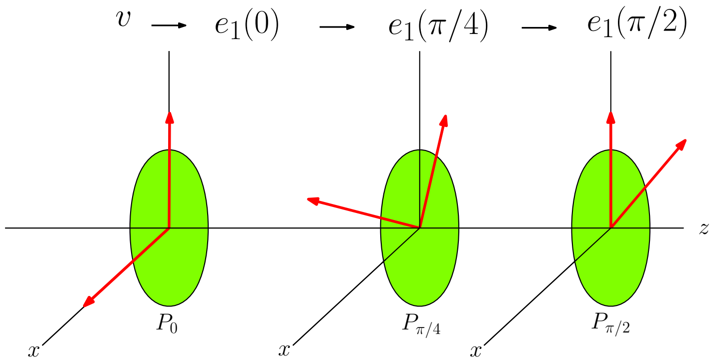
the probability of crossing the second polarizer $(e_1(0) \longmapsto e_1(\pi/2)\ )$ is equal to 0, as we had observed before.

The space of polarization states $\mathbb{C}^2$ has an inner product $\langle -, - \rangle$ defined as

$$\langle v, w \rangle = \langle (v_1, v_2), (w_1, w_2) \rangle = \overline{v_1} w_1 + \overline{v_2} w_2.$$

Hence, in general if v is the polarization of a photon, the probability of passing through $P(\theta)$ is equal to $|\langle v, e_1(\theta) \rangle|^2$. This is just elementary linear algebra: since $\{e_1(\theta), e_2(\theta)\}$ is orthonormal v can be written uniquely as $v = ae_1(\theta) + be_2(\theta)$. Thus, $|\langle v, e_1(\theta) \rangle|^2 = |\overline{a}|^2 = |a|^2$, is the probability of passing $P(\theta)$.

With this in mind, let us see again what happens when a photon encounters a row of three filters $P(0)$, $P(\pi/4)$ and $P(\pi/2)$


$v \longrightarrow e_1(0) \longrightarrow e_1(\pi/4) \longrightarrow e_1(\pi/2)$
z
P_0
$P_{\pi/4}$
$P_{\pi/2}$
x
x
x

Let $v = ae_1(0) + be_2(0)$ be its initial polarization. Passing $P(0)$ means v collapsing into $e_1(0)$. After this, the probability of also passing $P(\pi/4)$ would be

$$\begin{aligned} |\langle e_1(0), e_1(\pi/4) \rangle|^2 &= |\langle e_1,\ \cos(\pi/4)e_1 + \sin(\pi/4)e_2 \rangle|^2 \\ &= \left| \left\langle e_1, \frac{\sqrt{2}}{2} e_1 + \frac{\sqrt{2}}{2} e_2 \right\rangle \right|^2 = \frac{1}{2}. \end{aligned}$$

Consequently, in average, half of the photons that already passed $P(0)$ will also go through $P(\pi/4)$. After this second stage their polarizations will be equal to $e_1(\pi/4)$. Now, half of those photons will make it through $P(\pi/2)$, since

$$\begin{aligned}|\langle e_1(\pi/4),\ e_1(\pi/2)\rangle|^2 &= \left|\left\langle \frac{\sqrt{2}}{2}e_1 + \frac{\sqrt{2}}{2}e_2,\ \ \cos(\pi/2)e_1 + \sin(\pi/2)e_2\right\rangle\right|^2 \\ &= \left|\left\langle \frac{\sqrt{2}}{2}e_1 + \frac{\sqrt{2}}{2}e_2,\ e_2\right\rangle\right|^2 = \frac{1}{2}\end{aligned}$$

In consequence, we see that, on average, one eighth of the photon will cross the three filters.

1.9 Measurements and Hermitian operators

In classical mechanics functions in the phase state determine the measurable quantities of the system, like energy, momentum, etc. In quantum mechanics, on the other hand, *observables* are modeled by Hermitian operators.

An operator A in a Hilbert space $(\mathcal{H}, \langle -,-\rangle)$ is called *Hermitian* if A coincides with its adjoint A^*. That is, if $A^* = A$, where A^* is the unique operators satisfying the identity $\langle u, Av\rangle = \langle A^*v, u\rangle$, for all $u, v \in \mathcal{H}$.

Let us see how this formalism works in the case of the measurement of the polarization of a photon. Let T be the operator defined by the matrix (with respect to the standard basis e_1, e_2)

$$T = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

where $0 \le \theta \le 2\pi$ is such that

$$\begin{aligned} e_1(\theta) &= \cos\frac{\theta}{2}e_1 + \sin\frac{\theta}{2}e_2, \\ e_2(\theta) &= -\sin\frac{\theta}{2}e_1 + \cos\frac{\theta}{2}e_2. \end{aligned} \tag{1.9}$$

In the basis $B_\theta = \{e_1(\theta), e_2(\theta)\}$ the operator T can be written as

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

Therefore, $w_1 = e_1(\theta)$ and $w_2 = e_2(\theta)$ are the eigenvectors of T, with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$ respectively.

Suppose a photon is in a state $v = aw_1 + bw_2$. The probability that this measurement gives as a result the value 1 or -1 can be computed as $|a|^2 = |\langle v, w_1\rangle|^2$ and $|b|^2 = |\langle v, w_1\rangle|^2$, respectively. Therefore, the *expectation value* for such a measurement would be

$$\langle T\rangle_v = \lambda_1 |\langle v, w_1\rangle|^2 + \lambda_2 |\langle v, w_1\rangle|^2 .$$

In terms of T, the expectation value can be computed as $\langle v, Tv\rangle$. In fact:

$$\begin{aligned} \langle v, Tv\rangle &= \langle aw_1 + bw_2, aTw_1 + bTw_2\rangle \\ &= \langle aw_1 + bw_2, \lambda_1 a w_1 + \lambda_2 b w_2\rangle \\ &= \lambda_1 |a|^2 + \lambda_2 |b|^2 . \end{aligned}$$

In general, if T is a *complete* Hermitian operator in a complex Hilbert space $\mathcal{H}$ of dimension n, that is, if there is an orthonormal basis of eigenvectors w_n with eigenvalues λ_n for T, the expected value of the measurement determined by T when performed on a state v will be

$$\langle T \rangle_v = \langle v, Tv \rangle \,.$$

This is because the probability of obtaining λ_i would be $|c_i|^2 = |\langle v, w_i \rangle|^2$, where

$$v = c_1 w_1 + \cdots + c_n w_n.$$

Henceforth, the expected value is equal to

$$\langle T \rangle_v = \langle v, Tv \rangle = \left\langle \sum_{i=1}^{n} c_i w_i, \sum_{j=1}^{n} \lambda_j c_j w_j \right\rangle = \sum_{i=1}^{n} \lambda_i \bar{c}_i c_i = \sum_{i=1}^{n} \lambda_i \left|\langle v, w_i \rangle\right|^2 .$$

Intuitively this is the average value one obtains when performing the same experiment a large number of times on the same quantum state v.

1.10 Quantum Cryptography

The polarization states of a photon can be used to create a cryptography system that allows us to transmit information through a possibly non reliable channel ([Benn 2014]). This protocol, known as BB84, is used to pass on a binary sequence that, for example, may be used as a *key* in some other system of encryption, such as RSA ([RSA]).

Suppose Alex wants to send to Beatrice a binary sequence s of length n through an optic fiber. For this, we assume they both can use polarizer of types $P(0)$ and $P(\pi/4)$, as many as they need. This polarizers we will denote as + and ×, respectively. We also assume that Alex will send Beatrice no more than a photon at a time.

The protocol is a follows: Beforehand Alex chooses at random n polarizers of either type + and ×. Beatrice does the same, ignoring of course which sequence has been selected by Alex. Later she will use this sequence to register the information Alex will be sending to her.

To make things clear, suppose for instance that Alex wants to send Beatrice the sequence $s = \{1,\, 1,\, 0,\, 0,\, 0,\, 1\}$ and chooses polarizers $+\ \times\ \times\ +\ \times\ \times$. Suppose, on the other hand, that Beatrice chooses polarizers $+\ +\ \times\ +\ +\ \times$

$$\left|\begin{array}{ccccccc} \text{Alex} & + & \times & \times & + & \times & \times \\ \downarrow & & & & & & \\ \text{Beatrice} & + & + & \times & + & + & \times \end{array}\right| .$$

Alex send s using the following procedure: to transmit each of the digits of s he uses as a guide the polarizers he has chosen: Each time he wants to send a digit he chooses the corresponding polarizer $P(\theta)$, and prepares a photon in the direction of $e_1(\theta)$, if he wants to send a 1, and in the direction of $e_2(\theta)$ if he wants to send the digit 0. Once the photon reaches Beatrice, she uses as a filter the corresponding polarizer of her sequence.

In our example, Alex wants to send the first digit of the sequence, which is 1, and since his first polarizer is of type $P(0)$, he prepares a photon in the direction of $e_1(0)$. Beatrice, on the other hand, uses her first filter, $P(0)$, and observes that the photon goes through it registering the digit 1.

For the second digit, also a 1, Alex prepares a photon in the direction of $v_1 = 1/\sqrt{2}e_1 + 1/\sqrt{2}e_2$ (if he had wanted to send 0 he would have prepared a photon in the orthogonal direction $v_2 = -1/\sqrt{2}e_1 + 1/\sqrt{2}e_2$). Beatrice uses the second filter of her sequence, $P(0)$, and with probability 1/2 the photon will pass her filter. For the third digit of s, Alex should prepare a photon in the direction of v_2 and Beatrice should use the filter $P(\pi/4)$, and so on.

We observe that *each time Alex and Beatrice use the same type of polarizer the corresponding digit of the sequence is registered correctly. In case they use different filters, she will register the right digit only one half of the time.*

Once the protocol has ended they make public the complete list of polarizer used by each one of them, but without mentioning the digits that either one sent or received. Once Beatrice knows which polarizers Alex used, she proceeds to disregard those possibly faulty digits, i.e., those that correspond to the situation where both used different polarizers.

The following table summarizes the protocol. We have marked with $*$ the ignored digits in the sequence.

Alex	$+$	$\times$	$\times$	$+$	$\times$	$\times$
	1	1	0	0	0	1
	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
Beatrice	$+$	$+$	$\times$	$+$	$+$	$\times$
	1	$*$	0	0	$*$	1

We notice that in spite of the fact that everybody knows the polarizers used by Alex and Beatrice, no one, except Beatrice, could guess correctly the digit sent by Alex. After repeating this protocol a large number of times (changing the sequences of polarizers, but of course not s), Beatrice will eventually get all the digits of s correctly. This is because the probability of missing the same digit m times in a row is very small: $1/2^m$.

But suppose now that Eva is a spy who intercepts the channel. Then she would also get one half of the digits correctly. We just may imagine that she also fixes a random sequence of polarizers and does the same as Beatrice. However, after the photons go through Eva's polarizers only 3/4 of them will be transmitted as if no interference existed. This is because this occurs each time they choose the same polarizer, with probability 1/2 and when Eva chooses a different polarizer but gets by pure chance the correct digit. Thus last event occurs with probability 1/2 × 1/2. Thus, the interference of Eva in the transmission of *s will always introduce an error of* 25% *or greater* in the transmission of those digits that Alex and Beatrice know that are correct after the protocol is over. This is something that can certainly be discovered once the sequences of polarizers are made public. Thus, Eva may ruin the communication but she will never pass unnoticed.

The main point in the above discussion is the fact that *once a measurement is made there is no way to reconstruct the original state of the photon.* That information is lost for ever!

1.11 Quantum states

The quantum world turns out to be very counter-intuitive. To see why, let us analyze the celebrated interferometer experiment of Mach-Zehnder. The apparatus is constructed as shown below. A laser with frequency ω emits light that reaches a half-silvered mirror O, also known as a *beam splitter*. It is observed that exactly one half of the light is transmitted thorough the mirror and the other half is reflected in a perpendicular direction. Then both beams of light are directed by two ordinary mirrors towards another splitter at the right hand upper corner, which is parallel to the first one. The two beams are then mixed again at this second splitter. No light is observed at detector D_2 while D_1 registers light with intensity equal to that produced by the laser.

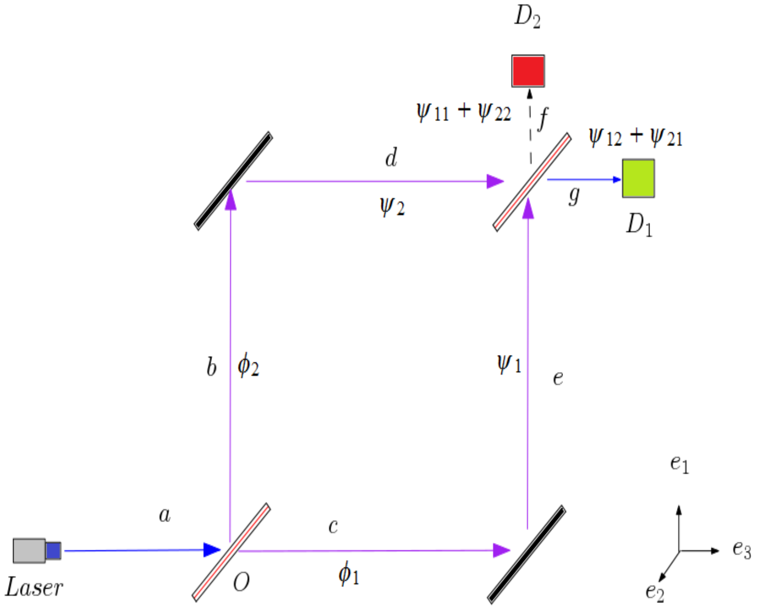


Figure 1.8: Mach-Zehnder experiment

Lets tray to understand what happens from a classical perspective. We consider the electric field on each constant plane z perpendicular to the direction of propagation of the light

$$E_z(t) = \sqrt{I}Re[e^{-\omega ti}(ae^{i\delta_1}e_1 + be^{i\delta_2}e_2)].$$

The effect of each splitter is to reduce to one half the intensity of the light that goes through it while maintaining its phase $e^{-i\omega t}$. The light that is reflected perpendicularly has also half the original intensity, but phase shifted by $\pi/2$. Hence, after crossing the first splitter the two fields will become

$$\begin{aligned}
\phi_1 &= \sqrt{\frac{I}{2}}Re\left[e^{-i\omega t}(ae^{i\delta_1}e_1 + be^{i\delta_2}e_2)\right]\\
\phi_2 &= \sqrt{\frac{I}{2}}Re\left[e^{\pi/2i}e^{-i\omega t}(ae^{i\delta_1}e_2 + be^{i\delta_2}e_3)\right]\\
&= \sqrt{\frac{I}{2}}Re\left[ie^{-i\omega t}(ae^{i\delta_1}e_2 + be^{i\delta_2}e_3)\right],
\end{aligned}$$

respectively. These can be written explicitly as:

$$\begin{aligned}
\phi_1 &= \frac{A}{\sqrt{2}}\cos(2\pi\omega t - \delta_1)e_1 + \frac{B}{\sqrt{2}}\cos(2\pi\omega t - \delta_2)e_2\\
\phi_2 &= \frac{A}{\sqrt{2}}\sin(2\pi\omega t - \delta_1)e_2 + \frac{B}{\sqrt{2}}\sin(2\pi\omega t - \delta_2)e_3.
\end{aligned}$$

It is also known that an ordinary mirror does not change neither the intensity nor the phase of the electromagnetic wave that it reflects. Hence, after ϕ_1 is reflected on the lower right mirror it only changes its direction, now perpendicular to e_1:

$$\psi_1 = \frac{A}{\sqrt{2}}\cos(2\pi\omega t - \delta_1)e_2 + \frac{B}{\sqrt{2}}\cos(2\pi\omega t - \delta_2)e_3.$$

Similarly, ϕ_2 after reflection changes to

$$\psi_2 = \frac{A}{\sqrt{2}}\sin(2\pi\omega t - \delta_1)e_1 + \frac{B}{\sqrt{2}}\sin(2\pi\omega t - \delta_2)e_2.$$

The last splitter will take ψ_1 and generate a beam ψ_{11} in its same direction, e_1, and another beam ψ_{12} in the perpendicular direction, e_3

$$\psi_{11} = \frac{A}{2}\cos(2\pi\omega t - \delta_1)e_2 + \frac{B}{2}\cos(2\pi\omega t - \delta_2)e_3,$$

$$\psi_{12} = \frac{A}{2}\sin(2\pi\omega t - \delta_1)e_1 + \frac{B}{2}\sin(2\pi\omega t - \delta_2)e_2.$$

Similarly, the splitter takes ψ_2 and splits it into ψ_{21} and ψ_{22} in the directions of e_3 and e_1, respectively

$$\psi_{21} = \frac{A}{2}\sin(2\pi\omega t - \delta_1)e_1 + \frac{B}{2}\sin(2\pi\omega t - \delta_2)e_2$$

$$\psi_{22} = \frac{-A}{2}\cos(2\pi\omega t - \delta_1)e_2 - \frac{B}{2}\cos(2\pi\omega t - \delta_2)e_3.$$

At the detector D_2 arrives the sum of the two waves $\psi_{11} + \psi_{22}$, which is equal to zero and *no light is registered there.* On the other hand, at D_1 we get $\psi_{12} + \psi_{21}$, a beam of light with the same intensity I as the original beam coming from the laser, but with phase shifted $\pi/2$.

This seems to explain everything until a detector D' at the first splitter is placed. This detector witnesses the passing of a photon through the splitter one half of the times on average, as if each photon either passes or is reflected with probability 1/2. But then a surprising situation occurs: Detector D_2 starts receiving light! In fact, the presence of the detector D' makes that both D_1 and D_2 detect light with intensity $I/2$.

The classical corpuscular model of light will not explain the results of this experiment either. If we activate D' then everything makes sense since each photon that reaches the splitter either goes through it or is reflected perpendicularly with equal probability, then only one half of the photons that pass the first splitter will pass the second while the other half will be reflected. Hence 1/4 of the original photons will go through the first splitter and are then reflected at the second splitter. Similarly 1/4 of the photons will be reflected at the first splitter and will pass the second one. This accounts for 1/2 of all the photons emitted by the laser to be found at detector D_1 and the other half at D_2. However, when D' is turned off all the photons end up at D_1 which is absurd! Somehow the presence of the detector alters the outcome of the experiment. As we shall see later, what changes the outcome is not any sort of physical interaction with D' but the fact that we have *gained information* about its trajectory. This is a truly remarkable fact!

1.12 Quantum Explanation

In the quantum model we can explains what happens. The first splitter has the effect of putting the state of the photon in a superposition of states $\phi = \frac{1}{\sqrt{2}}(|c\rangle + i\,|b\rangle)$. Here $\{|c\rangle,\, |b\rangle\}$ denote an arbitrary orthonormal base of $\mathbb{C}^2$ that we relate to the two possible paths for each photon. The factor i is chosen to take into consideration the phase change $e^{\pi/2i} = i$. By placing and turning on D' we make the state ϕ *collapse* into either $|c\rangle$ or $|b\rangle$ depending of which trajectory is observed: The transition $\phi \to |c\rangle$ means that it takes route c while the transition $\phi \to |b\rangle$ means that it has taken route b. On the other hand, the effect of the two ordinary mirrors is to transform $|c\rangle$ into $|e\rangle$ and $|b\rangle$ into $|d\rangle$.

If no measurement is done at either of the two splitters the photon's state will evolve as follows: $\phi = \frac{1}{\sqrt{2}}(|c\rangle + i\,|b\rangle)$ will evolve into $\phi' = \frac{1}{\sqrt{2}}(|e\rangle + i\,|d\rangle)$. At the second splitter $|e\rangle$ is transformed into $\frac{1}{\sqrt{2}}(|f\rangle + \frac{i}{\sqrt{2}}\,|g\rangle)$, and $|d\rangle$ into $\frac{1}{\sqrt{2}}(|g\rangle + \frac{i}{\sqrt{2}}\,|f\rangle)$. Henceforth, ϕ' is transformed into

$$\begin{aligned}\phi_2 \to \frac{1}{\sqrt{2}}\left[\frac{1}{\sqrt{2}}\,|f\rangle + i\frac{1}{\sqrt{2}}\,|g\rangle + i(\frac{1}{\sqrt{2}}\,|g\rangle + i\frac{1}{\sqrt{2}}\,|f\rangle)\right] \\ = \frac{1}{2}\,|f\rangle + \frac{i}{2}\,|g\rangle + \frac{i}{2}\,|g\rangle - \frac{1}{2}\,|f\rangle = i\,|g\rangle\,.\end{aligned}$$

Thus, the probability of being detected at D_1 would be 1, and zero at D_2, as observed.

When D' is activated, ϕ collapses into either $|c\rangle$ or $|b\rangle$. Consequently, at the second splitter its state will be either $|e\rangle$ or $|d\rangle$. This state is then transformed into either $\frac{1}{\sqrt{2}}\,|f\rangle + \frac{i}{\sqrt{2}}\,|g\rangle$ or $\frac{1}{\sqrt{2}}\,|g\rangle + \frac{i}{\sqrt{2}}\,|f\rangle$. Measurement at D_1 collapses the states into $|f\rangle$ (not detected) or $|g\rangle$ (detected) with equal probability. Thus, we observe light at both detectors D_1 and D_2 with equal intensity.

1.13 The Elitzur-Vaidman experiment

The Mach-Zender experiment has a curious interpretation due to the Israeli physicists Avshalom Elitzur and Lev Vaidman. Suppose that the mirror located at the low right corner of the interferometer is attached to a certain type of bomb. These bombs come with a photodetector so sensitive that is activated by the presence of even one photon. The photocell, when activated, sends a signal that makes the bomb explode. We could think of each bomb as a photon detector, one quite dramatic though.

Many bombs, however, come with a defect that prevents the mechanism to be set in action so that when a photon hits their mirror, it is just reflected, but the bomb does not explode. A defective bomb, on the other hand, would just behave as the ordinary mirror in the Mach-Zender apparatus.

For a defective bomb the original state of the photon $\phi = \frac{1}{\sqrt{2}}(|c\rangle + i\,|b\rangle)$ never collapses and the photon ends up at D_1. It is never detected at D_2. A good bomb, however, will explode one half of the times on average, each time ϕ reaches its mirror and collapses into $|c\rangle$. If ϕ collapses into $|b\rangle$, the photon will be reflected and the bomb will not explode. After reflection, the photon will end up at either D_1 or D_2 with equal probability.

The surprising quantum behavior of photons implies that for a large batch of bombs about one thirds can be certified as good bombs without destroying them! In fact, suppose we test a very large batch of n bombs using them as mirror detectors in the interferometer. On average half of the good bombs will explode in the process. For the other half, detectors D_1 and D_2 are activated with equal probability.

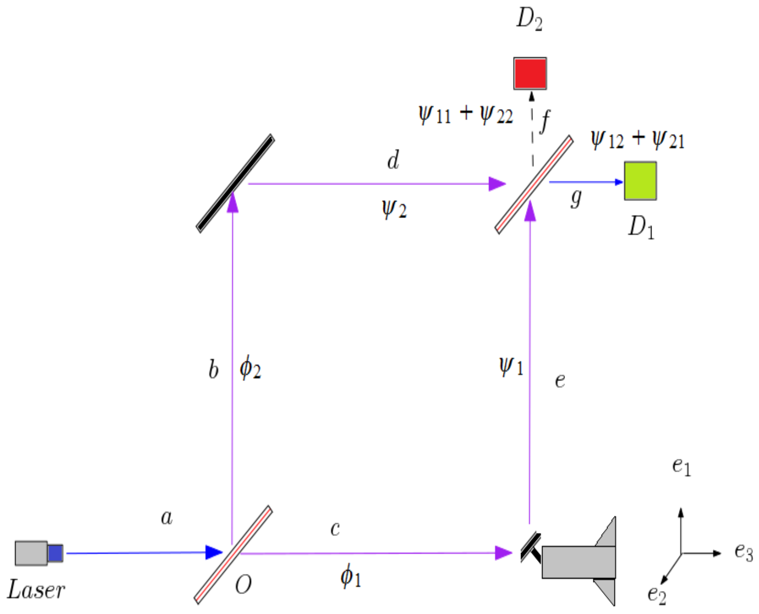


Figure 1.9: Bomb detector

We can be sure that those bombs that activate D_2 are necessarily good bombs, since a defective bomb never makes the photon state collapse, and consequently every photon ends up at D_1. Then, on average, $1/4$ of all good bombs will end up activating D_2, and can be rescued. Moreover, the bombs that activated D_1 can be retested, and from this new batch we can save $1/4 \times 1/4$ of all the good ones. Continuing in this way at the end we could rescue $1/4+1/16+1/64+\cdots=1/3$ of all good bombs.

Chapter 2

Axioms of Quantum Mechanics

2.1 Introduction

The examples we discussed in the first chapter can be formulated within a general formal framework. An isolated quantum system corresponds to a set of states or vectors of a certain complex Hilbert space H endowed with a Hermitian product, which we will denote by $\langle -,-\rangle$. Two vectors ϕ and ψ represent the same physical state if $\phi = \alpha\psi$, for a non-zero complex α. Hence, states are usually taken as unit vectors, after they are normalized. In the first three chapters of these notes, all spaces will be finite-dimensional. We will adopt the physicists' convention that the Hermitian product is a bilinear function, linear conjugate in the first entry, and linear in the second. That is, it satisfies:

1. $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$
2. $\langle v, w_1 + w_2\rangle = \langle v, w_1\rangle + \langle v, w_2\rangle$
3. $\langle \alpha v, w\rangle = \overline{\alpha}\langle v, w\rangle$, and $\langle v, \alpha w\rangle = \alpha\langle v, w\rangle$.
4. $\langle v, w\rangle = \overline{\langle w, v\rangle}$
5. $\langle v, v\rangle \geq 0$, and it is equal to zero only if $v = 0$

We will use interchangeably the notation of mathematicians and physicists (Dirac's notation): the vectors will be denoted by $v, \phi, f...$, or by $|v\rangle$, $|\phi\rangle$, $|f\rangle$, ...The functional induced by the inner product with a fixed vector v will be denoted by $\langle v, -\rangle$ or as $\langle v|$. This last symbol is understood as the functional one that takes vectors w and produces $\langle v, w\rangle$, which physicists usually denote by $\langle v|w\rangle$. For example, if $\{e_i : i = 1, \ldots n\}$ is an orthonormal basis for H, in Dirac's notation the operator T that sends $e_i \to \lambda_i e_i$ is written as

$$T = \sum \lambda_i |e_i\rangle \langle e_i| .$$

The action of T on a vector $|v\rangle$ would be:

$$T(|v\rangle) = \sum \lambda_i |e_i\rangle \langle e_i|v\rangle = \sum \alpha_i\lambda_i |e_i\rangle ,$$

where $|v\rangle = \sum \alpha_i |e_i\rangle$, since

$$\langle e_i|v\rangle = \sum \alpha_j \langle e_i|e_j\rangle = \alpha_j.$$

The generators of the tensor product $v_i \otimes w_j$ of two Hilbert spaces H and H' are usually denoted by $|v_i, w_i\rangle$. In many situations when referring to an explicit orthonormal basis $\{e_i : i = 1, \ldots n\}$ the vectors e_i are simply denoted by $|i\rangle$. Also, $e_i \otimes e_j$ is denoted by $|i, j\rangle$, an triple tensor products $e_i \otimes e_j \otimes e_k$ by $|i, j, k\rangle$....

The *observables* in H correspond to *Hermitian* operators (we recall A is called Hermitian if $\langle v, Aw\rangle = \langle Av, w\rangle$). With respect to an orthonormal basis $e = \{e_i\}$ an operator A is Hermitian if and only if $M^* = M$, where M^* denotes the conjugate adjoint matrix of the matrix M representing the operator A in that basis.

When A has n different eigenvalues, the formalism should be understood as follows: in the case in which $e = \{e_i\}$ is the set of eigenvectors of A, with corresponding eigenvalues λ_i then the measurement associated with A on a state vector ϕ produces one of the possible values λ_i. Each value is obtained when the vector ϕ collapses in the state e_i, which occurs (this is the fundamental axiom) with probability

$$|\alpha_i|^2 = |\langle e_i|\phi\rangle|^2 = \frac{|\langle e_i|A\phi\rangle|^2}{\lambda_i^2},$$

where $\phi = \sum \alpha_i e_i$ is the expansion of ϕ in the base e.

The Spectral Theorem guarantees that every Hermitian operator is diagonalizable with real eigenvalues. This means that there exists a set of subspaces E_i orthogonal to each other ($E_i \perp E_j$, if $i \neq j$) such that $H = E_1 \oplus \cdots \oplus E_k$. Associated with each E_i there is a real eigenvalue λ_i, that is: $Tv = \lambda_i v$, for each $v \in E_i$. The operator T is called *complete or non-degenerate*, if E_i is the subspace of dimension one generated by a single eigenvector e_i.

If T is non-degenerate, we had seen that $|\langle e_i, \phi\rangle|^2$ is to be interpreted in the quantum formalism as the probability that the state represented by ϕ would collapse into e_i, which corresponds to "having measured λ_i". In general, for a not necessarily complete T, we can write $H = E_i \oplus H'$, where H' is the orthogonal complement of E_i, then a fundamental axiom tells us that the probability of measuring λ_i is given by $|\phi_i|^2$, where $\phi = \phi_i + \phi\prime$ is the decomposition of ϕ in this direct sum. The measurement will then collapse ϕ into the new state $\phi_i / |\phi_i|$ (we must normalize ϕ_i to obtain a unit vector). This collapse is in turn associated with the projection operator on E_i, $P(\phi) = \phi_i$.

2.2 Evolution of a quantum state

Recall that a linear operator U is called unitary if $\langle Uv, Uw\rangle = \langle v, w\rangle$. This condition is equivalent to saying that $UU^* = U^*U = I$, (here I denotes the identity matrix). The third fundamental axiom tells us that the evolution in time of a state ϕ is given by a family of unitary operators $U(s,t)$. If we assume $\phi(0) = \phi$, then $\phi(s) = U(s,t)\phi(t)$. The latter forces this family of operators to satisfy the so-called *group property*: $U(s,t')U(t',t) = U(s,t)$ and $U(t,t) = I$. Let us now see that the existence of $U(s,t)$ implies the famous Schrödinger evolution equation.

For fixed t define the operator $R(s) = U\ (s,t)$. In a finite dimensional space it is easy to show that R admits a Taylor series expansion

$$R(t+\delta) = R(t) + \delta R^{(1)}(t) + \frac{R^{(2)}(t)}{2!}\delta^2 + \cdots$$

since each entry in any matrix representing R has a Taylor's series. The operators $R^{(i)}(t)$ would then correspond to taking the i-th derivative of each entry. Since $R(t) = U(t,t) = I$, we obtain

$$R(t+\delta) - I = \delta R^{(1)}(t) + \delta^2 B(t).$$

On the other hand,

$$\begin{aligned}\frac{1}{\delta}\left[\phi(t+\delta)-\phi(t)\right] &= \frac{1}{\delta}\left[U(t+\delta,t)\phi(t)-\phi(t)\right]\\ &= \frac{1}{\delta}\left[R(t+\delta)-I\right]\phi(t).\end{aligned}$$

Taking limits when $\delta \to 0$ we see that

$$\phi'(t) = R^{(1)}(t)\phi(t).$$

If we multiply both sides by $i\hbar$, ($\hbar = h/(2\pi)$ is Planck's constant divided by 2π), and setting $H = i\hbar R^{(1)}(t)$ one obtains

$$i\hbar\ \phi'(t) = H(t)\phi(t).$$

The purpose of multiplying both sides by the factor $i\hbar$ is to change $R^{(1)}(t)$ by a Hermitian operator with units of energy. The operator H is in fact Hermitian: Since $I = R(t+\delta)R^*(t+\delta)$ by expanding in Taylor's series one obtains

$$\begin{aligned}I = R(t+\delta)R^*(t+\delta) &= \left[I - \frac{i}{\hbar}\delta H(t) + \delta^2 B(t)\right]\left[I + \frac{i}{\hbar}\delta H^*(t) + \delta^2 B^*(t)\right]\\ &= I + \frac{i\delta}{\hbar}(H^*(t)-H(t)) + \delta^2 C(t).\end{aligned}$$

Henceforth

$$\frac{i\delta}{\hbar}(H^*(t)-H(t)) + \delta^2 C(t) = 0.$$

After dividing by δ and taking the limits as $\delta \to 0$ we see that $H^*(t) = H(t)$.

The operator H represents the *observable that correspond to the energy of the system* and is called the *Hamiltonian*.

As an example, let us consider a monochromatic beam of light, a Laser, of *angular* frequency ω, that is $\omega = \nu/(2\pi)$ where ν is the frequency of the Laser. Suppose it consists of photons linearly polarized in the direction of some vector $v \in \mathbb{C}^2$. For simplicity we will assume that $v = e_1$. If there is no further measurement, at any time t the state of polarization of the photon is given by (1.7):

$$\phi(t) = e_1. \tag{2.1}$$

On the other hand, the dynamics of the system is determined by the equation

$$i\hbar\phi(t)' = H(t)\phi(t), \tag{2.2}$$

where H is the associated Hamiltonian. In this case one sees that by taking

$$H = \begin{bmatrix} -\hbar\omega & 0 \\ 0 & \hbar\omega \end{bmatrix}$$

the solution of (2.2) is the function $\phi(t) = e_1$.

In fact, if $\phi(t) = \alpha(t)e_1 + \beta(t)e_2$ then (2.2) is equivalent to the system of differential equations

$$\alpha'(t) = \frac{i}{\hbar}\hbar\omega\ \alpha(t),\ \ \beta'(t) = \frac{-i}{\hbar}\hbar\omega\ \beta(t).$$

This system has as solution

$$\alpha(t) = \alpha_0 e^{i\omega t},\ \beta(t) = \beta_0 e^{-i\omega t},$$

where α_0 and β_0 are determined by the initial conditions $\phi(0) = e_1$. Hence, $\alpha_0 = 1$, $\beta_0 = 0$, and for all $t \geq 0$ the state of the photon is given by $\phi(t) = e^{i\omega t} e_1$, which is the same state as in (2.1).

The Hamiltonian in this case represents the two possible levels of energy of the photon $E = \hbar\omega$ and $E = -\hbar\omega$.

A quantum system is called *insulated* if its Hamiltonian is constant, i.e., if $H(t) = H$. In this particular case the equation $i\hbar\, \phi(t)' = H\phi(t)$ can be solved directly as

$$\phi(s) = \exp\left(-\frac{i(s-t)}{\hbar} H\right) \phi(t). \tag{2.3}$$

The term $\exp(-)$ denotes the exponential function of an operator. If A is a matrix representation of an operator, its exponential is defined as

$$\exp(A) = \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

From (2.3) one can compute explicitly the family of operators $U(s,t)$ as

$$U(s,t) = \exp\left(-\frac{i(s-t)}{\hbar} H\right).$$

Summarizing the discussion above we may enunciate the following axioms for quantum mechanics.

Axiom 2.2.1. A quantum system is modeled by a complex vector Hilbert space $(H, \langle -,- \rangle)$. Two vector ϕ and ψ describe the same state if $\phi = \alpha\psi$, for a complex number $\alpha \neq 0$. Henceforth, it is customary to represent any state by a unitary vector in its class.

Axiom 2.2.2. The quantum system that results of considering several interacting quantum systems represented by Hilbert spaces $(H_i, \langle -,- \rangle_i)$ is modeled by the tensor product space: $H = H_1 \otimes \cdots \otimes H_n$. The inner product in H is given by $\langle -,- \rangle = \langle -,- \rangle_1 \cdots \langle -,- \rangle_n$

Axiom 2.2.3. An observable is identified with a Hermitian operator H. If $(\lambda_i,\, E_i)$ are the eigenvalues and eigenspaces of H, then to perform a measurement on ϕ with end result λ_i means that ϕ is transformed "collapses" into $\psi_i / |\psi_i|$, where ψ_i is the projection of ϕ on E_i. The probability of this event is given by $|\psi_i|^2$. We notice that since $H = E_1 \oplus \cdots \oplus E_r$ one can write $\phi = \psi_1 + \cdots + \psi_r$ where

$$1 = |\phi|^2 = |\psi_1|^2 + \cdots + |\psi_r|^2$$

When H has n different eigenvalues with eigenvectors $\{e_i\}$, the probability of obtaining λ_i (and thus collapsing into e_i) can be computed as

$$|\alpha_i|^2 = |\langle e_i, \phi \rangle|^2,$$

where $\phi = \sum \alpha_i e_i$ is the expansion of ϕ in the orthonormal base $\{e_i\}$.

Axiom 2.2.4. The evolution of a state ϕ is given by a family of unitary operators $U(s,t)$ satisfying:

a $\phi(s) = U(s,t)\phi(t)$ $(\phi(0) = \phi)$.

b If $t \leq t' \leq s$ then $U(s,t')U(t',t) = U(s,t)$

c $U(t,t) = I$.

From this we deduce the equation of evolution or Schrödinger (wave) equation:

$$i\hbar\phi'(t) = H(t)\phi(t),$$

where $H(t)$ is the Hamiltonian operator of the system that corresponds to the energy observable.

2.3 Quantum entanglement

Axiom 2.2.2 describing several interacting quantum systems requires a detailed explanation. According to this axiom, two photons p_1 and p_2 are not regarded as separated entities but as one single mathematical vector. For instance, the polarization state of the pair (p_1, p_2) is not the pair of vectors (v_1, v_2) where v_i denotes the polarization state of p_i. In general, the polarization state of the system will be described by a unitary vector in $\mathbb{C}^2 \otimes \mathbb{C}^2$. In the standard basis of $\mathbb{C}^2$ one can write this general state v as

$$v = \alpha_{11}\ e_1 \otimes e_1 + \alpha_{12}\ e_1 \otimes e_2 + \alpha_{21}\ e_2 \otimes e_1 + \alpha_{22}\ e_2 \otimes e_2,$$

where $\sum\limits_{i,j=1,2.} |\alpha_{ij}|^2 = 1$.

We notice that the polarization of the system formed by the two photons is in general not equal to $v_1 \otimes v_2$ either. For instance, in some radioactive cascades a pair of photons that move in opposite directions is produced. By conservation of momentum and parity, the polarization state of this system turns out to be

$$\eta = \frac{1}{\sqrt{2}}(e_1 \otimes e_1 + e_2 \otimes e_2).$$

It is easy to see, however, that there are no pair of vectors w_1, w_2 in $\mathbb{C}^2$ satisfying $\eta = w_1 \otimes w_2$. This particular state is called an *entangled state.*

The idea of an entanglement of two particles was considered perturbing by Einstein. In his opinion, this seemed to imply a violation of the locality principle. That is, by manipulating one of the particle you seemed to be able to modify the state of the other, no matter how far apart they could be. But no signal can travel faster than light, which seemed to be a contradiction. To prove his point, Einstein suggested a thought experiment famously known as EPR, which are the initial of three eminent physicists, Albert Einstein, Boris Podolsky and Nathan Rosen.

In the EPR experiment two photons are emitted in the entangle state η described above. Separated by several light years two observers (let's called them Alex and Beatrice) measure the polarization of each one of the photons in some particular direction that they choose randomly once the photon are in full fly. Then, each one of them will record the passing of the photon or if it is absorbed by the chosen filter. If Alex and Beatrice equally measure either a passing or an absorption, we call the result of the experiment a *coincidence.* Otherwise, we call it a *discrepancy.*

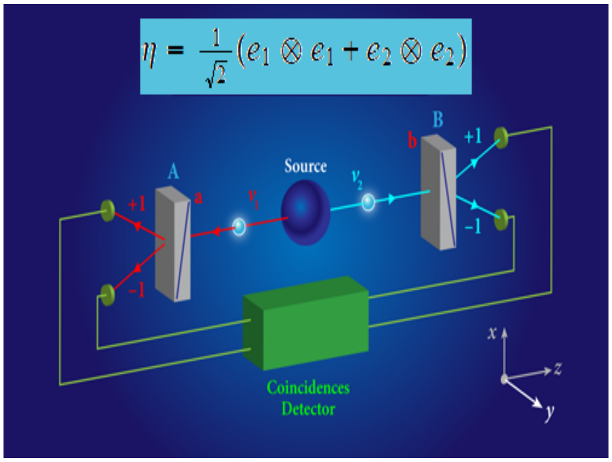

$\eta = \frac{1}{\sqrt{2}}(e_1 \otimes e_1 + e_2 \otimes e_2)$
A
a
+1
-1
v_1
Source
v_2
B
b
+1
-1
Coincidences
Detector
x
z
y

Figure 2.1: EPR experiment

It is experimentally observed that

(*) *A coincidence is always obtained if both Alex and Beatrice use the same filter.*

We notice that this fact rules out the possibility of modeling the polarization of the two photons as an independent pair of polarizations of each one separately, (v_1, v_2). This is because if Alex and Beatrice choose the same type of filter, it may happen that Alex registers his photon passing while Beatrice would see it being absorbed, or the other way around. Notice that even if $v_1 = v_2$, let's say $v_1 = \cos(\alpha)e_1 + \sin(\alpha)e_2$, then there is a 100% certainty of obtaining a coincidence only if Alex and Beatrice happen to choose filter $P(\alpha)$. For any other choice of filters this will occur, only with some probability, contradicting the empirical evidence. We then see that, apparently, the only way for (*) to hold would be if we assume that in these type of radioactive cascades *each of the photons that are produce already carry a well determined set of polarizations*, one in each possible direction $0 \leq \alpha < 2\pi$. There is no other way one of the photons could influence its twin instantly, since Special Relativity prohibits a causality relation between them. The situation is very much as in our toy model for polarization, but in this case we leave open the possibility that the polarization of each photon may change once a measurement is performed.

To simplify matters, suppose Alex and Beatrice decide to use only three types of filters, $A = P(0), B = P(120)$ and $C = P(240)$. Then, once both photons detach, they must carry a equally well defined set of polarizations in each one of these three directions, a situation we represent by using a $+$ and $-$ sign. For instance, we would say that they are both of type $\overset{+}{A}\overset{-}{B}\overset{+}{C}$ if they would pass filters A and C, but not B. Notice that even though Alex and Beatrice do not know in advanced what they will measure, the outcome of the experiment, nevertheless, is already determined, but Alex and Beatrice happen to ignore it.

Let's try to see in this situation how likely it is that a coincidence occurs. The next table shows for each one of the eight possibilities for the polarizations the number of coincidences that

Alex and Beatrice will observe.

A	B	C	A	B	C	Number of coincidences
$+$	$+$	$+$	$+$	$+$	$+$	9 out of 9
$+$	$+$	$-$	$+$	$+$	$-$	5 out of 9
$+$	$-$	$+$	$+$	$-$	$+$	5 out of 9
$+$	$-$	$-$	$+$	$-$	$-$	5 out of 9
$-$	$+$	$+$	$-$	$+$	$+$	5 out of 9
$-$	$+$	$-$	$-$	$+$	$-$	5 out of 9
$-$	$-$	$+$	$-$	$-$	$+$	5 out of 9
$-$	$-$	$-$	$-$	$-$	$-$	9 out of 9

We then see that the probability of obtaining a coincidence would be:

$$\begin{aligned} & P = 1/8 \times (9/9 + 5/9 + 5/9 + 5/9 + 5/9 + 5/9 + 5/9 + 9/9) \\ = & \; 1/8 \times 48/9 = 2/3. \end{aligned}$$

But the experiment shows a probability of 1/2 for either a coincidence or a discrepancy! Henceforth, the classical explanation cannot be correct!

2.4 Analysis of the EPR experiment

Let us denote $E(\alpha) = \{e_1(\alpha), e_2(\alpha)\}$ and $E(\beta) = \{e_1(\beta), e_2(\beta)\}$ two bases corresponding to filters $P(\alpha)$ and $P(\beta)$. The standard basis can be express in either of these bases as

$$\begin{aligned} e_1 &= \cos(\alpha)e_1(\alpha) - \sin(\alpha)e_2(\alpha), \\ e_2 &= \sin(\alpha)e_1(\alpha) + \cos(\alpha)e_2(\alpha), \end{aligned}$$

and similarly for β. Thus, the state η can be written as

$$\begin{aligned} \eta = & \tfrac{1}{\sqrt{2}}[\cos(\alpha)e_1(\alpha) - \sin(\alpha)e_2(\alpha)] \otimes [\cos(\beta)e_1(\beta) - \sin(\beta)e_2(\beta)] \\ & + \tfrac{1}{\sqrt{2}}[(\sin(\alpha)e_1(\alpha) + \cos(\alpha)e_2(\alpha)] \otimes [\sin(\beta)e_1(\beta) + \cos(\beta)e_2(\beta)] \end{aligned}$$

$$\begin{aligned} = & \frac{1}{\sqrt{2}}\cos(\alpha-\beta)e_1(\alpha) \otimes e_1(\beta) + \frac{1}{\sqrt{2}}\cos(\alpha-\beta)e_2(\alpha) \otimes e_2(\beta) + \\ + & \frac{1}{\sqrt{2}}\sin(\alpha-\beta)e_1(\alpha) \otimes e_2(\beta) - \frac{1}{\sqrt{2}}\sin(\alpha-\beta)e_2(\alpha) \otimes e_1(\beta)), \end{aligned} \tag{2.4}$$

If Alex and Beatrice used polarizers $P(\alpha)$ and $P(\beta)$, the state η would collapse into one of the four possible states $e_i(\alpha) \otimes e_j(\beta)$, $i, j = 1, 2$ with probability equal to the norm of the corresponding coefficient squared. For instance, the probability of measuring $e_1(\alpha) \otimes e_1(\beta)$ (p_1 and p_2 pass filters $P(\alpha)$ and $P(\beta)$, respectively) would be $\frac{1}{2}\cos^2(\alpha-\beta)$. In a similar fashion, the probability of p_1 passing and p_2 being absorbed would be: $\frac{1}{2}\sin^2(\alpha-\beta)$. The probability of observing a discrepancy would be $\sin^2(\alpha-\beta)$. Hence, the quantum model explains (*): When using the same filter they will always observe a coincidence.

Suppose now that they use three filters as above. If it happens that Alex registers first the passing of his photon, let us compute the possibility that Beatrice does the same. This sequence of events amounts to the quantum state first collapsing into (Alex measurement)

$$\eta_1 = \cos(\alpha-\beta)e_1(\alpha)\otimes e_1(\beta) + \sin(\alpha-\beta)e_1(\alpha)\otimes e_2(\beta).$$

And then η_1 collapsing into $e_1(\alpha)\otimes e_1(\beta)$, something that occurs with probability $\cos^2(\alpha-\beta)$. This number equals 1 if $\alpha=\beta$ and 1/4 if $\alpha\neq\beta$, for α and β two angles in the set $\{0, 120°, 240°\}$ Henceforth, if Alex registers that his photon passed, the probability of observing a coincidence would be

$$P = \frac{1}{3}\times 1 + \frac{1}{3}\times\frac{1}{4} + \frac{1}{3}\times\frac{1}{4} = \frac{1}{2}.$$

The formalism is consistent and predicts that the end result will not depend on who performs the measurement first. If, for instance, Beatrice records the non passing of her photon, the probability of Alex recording the same would be the probability of the following sequence of collapses:

$$\begin{aligned} \eta &\rightarrow \eta_2 = \cos(\alpha-\beta)e_2(\alpha)\otimes e_2(\beta) + \sin(\alpha-\beta)e_1(\alpha)\otimes e_2(\beta) \\ \eta_2 &\rightarrow e_2(\alpha)\otimes e_2(\beta) \end{aligned}$$

The probability of this event is $1/2\times\cos^2(\alpha-\beta)$ since the probability of the first event is 1/2. The result is the same as if we had measured both polarizations at the same time and had ended up with both photons being absorbed, the state $e_2(\alpha)\otimes e_2(\beta)$.

2.5 Bell's Inequality

Our toy model for polarization could be enhanced to a general classical model of local type. This idea was suggested by Einstein himself, and developed by J. Bell and other famous physicists. Let us see why this more general local model cannot hold either.

Let us suppose that every photon in nature comes with a fixed polarization in each direction. More precisely, we assume that all photons are distributed according to certain probability density function $\rho(\lambda)$, $\lambda\in[0,2\pi]$

$$\rho(\lambda)\geq 0, \int_0^{2\pi}\rho(\lambda)d\lambda = 1.$$

We also assume that there exists a function $A(\lambda,-)$ taking values ±1, such that $A(\lambda,\alpha)=+1$, if every photon of type p_λ passes a filter $P(\alpha)$, and $A(\lambda,\alpha)=-1$, if it is absorbed. Hence, the function $B(\lambda,\alpha)=\frac{1}{2}(A(\lambda,\alpha)+1)$ takes values 1 and 0 in each of the above cases.

In terms of the function $A(\lambda,\alpha)$, the probability that a randomly chosen photon passes $P(\alpha)$ (register $+1$) would be given by

$$P = \int_0^{2\pi} B(\lambda,\alpha)\rho(\lambda)d\lambda.$$

Bell's idea consists in measuring different instances of the expected value $P(\alpha,\beta)$ of the correlation function, defined as $C(\alpha,\beta,\lambda)=A(\lambda,\alpha)A(\lambda,\beta)$, for pairs of photons in the EPR experiment. Notice that Alex and Beatrice observe a coincidence if $C(\alpha,\beta,\lambda)=+1$; and $C(\alpha,\beta,\lambda)=-1$, if they observe a discrepancy.

The expected value of the correlation function is then given by

$$P(\alpha,\beta)=\int_0^{2\pi}\rho(\lambda)A(\lambda,\alpha)A(\lambda,\beta)d\lambda.$$

Let us estimate $P(\alpha,\beta)+P(\alpha,\gamma)$ for three possible directions α,β,γ:

$$\begin{aligned}P(\alpha,\beta)+P(\alpha,\gamma)&=\int_0^{2\pi}\rho(\lambda)[A(\lambda,\alpha)A(\lambda,\beta)+A(\lambda,\alpha)A(\lambda,\gamma)]d\lambda\\&=\int_0^{2\pi}\rho(\lambda)[A(\lambda,\alpha)A(\lambda,\beta)+A(\lambda,\alpha)A(\lambda,\gamma)A(\lambda,\beta)^2]d\lambda\ (\text{since } A(\lambda,\beta)^2=1)\\&=\int_0^{2\pi}\rho(\lambda)A(\lambda,\alpha)A(\lambda,\beta)[1+A(\lambda,\gamma)A(\lambda,\beta)]d\lambda.\end{aligned}$$

But, $A(\lambda,\alpha)A(\lambda,\beta)\leq 1$, and consequently

$$\begin{aligned}P(\alpha,\beta)+P(\alpha,\gamma)=\int_0^{2\pi}&\rho(\lambda)A(\lambda,\alpha)A(\lambda,\beta)[1+A(\lambda,\gamma)A(\lambda,\beta)]d\lambda\leq\\&\int_0^{2\pi}\rho(\lambda)[1+A(\lambda,\gamma)A(\lambda,\beta)]d\lambda\\&\leq\int_0^{2\pi}\rho(\lambda)d\lambda+\int_0^{2\pi}\rho(\lambda)A(\lambda,\gamma)A(\lambda,\beta)]d\lambda\\&\leq 1+P(\beta,\gamma).\end{aligned}$$

Summarizing our discussion

$$P(\alpha,\beta)+P(\alpha,\gamma)\leq 1+P(\beta,\gamma).\qquad\text{(Bell's inequality)}$$

On the other hand, we already know that the probability of recording a coincidence is equal to

$$\frac{1}{2}\cos^2(\alpha-\beta)+\frac{1}{2}\cos^2(\alpha-\beta)=\cos^2(\alpha-\beta).$$

Similarly, the probability of recording a discrepancy is equal to $\sin^2(\alpha-\beta)$. Therefore, $P(\alpha,\beta)$ must be equal to

$$P(\alpha,\beta)=\cos^2(\alpha-\beta)-\sin^2(\alpha-\beta)=\cos 2(\alpha-\beta).$$

Taking $\alpha=0$, $\gamma=\beta+\pi/4$ we see that

$$\begin{aligned}&P(\alpha,\beta)=\cos(2\beta)\\&P(\alpha,\gamma)=\cos(2\gamma)=\cos(2\beta+\pi/2)=-\sin(2\beta)\\&P(\beta,\gamma)=\cos(\pi/2)=0\end{aligned}$$

This would imply

$$\cos(2\beta)-\sin(2\beta)\leq 1+0.$$

But if we take $\beta=7\pi/8$ one has

$$\cos(2\beta)-\sin(2\beta)=\cos(7\pi/4)-\sin(7\pi/4)=\sqrt{2}.$$

Hence, Bell's inequality cannot hold. The experiments of Alain Aspect and many other that came after him vindicate the quantum model [AD].

2.6 Double-slit experiment

In this section we revisit the double-slit experiment. Let us recall that a beam of monochromatic light is passed through two slits separated by a very small distance. The light passing through the slits is then projected onto a screen, on which a series of interference fringes are observed, a phenomenon explained by Young in the 18th century using the wave model of light.

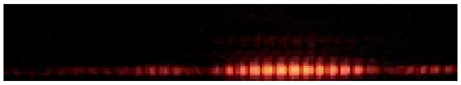


Figure 2.2: Interference pattern

However, as we already ramarked, this phenomenon would be impossible if we think of a beam of light as a beam composed of small individual photons. In the corpuscular model each photon should take one and only one of two possible trajectories: passing through the upper slit or passing through the lower slit.

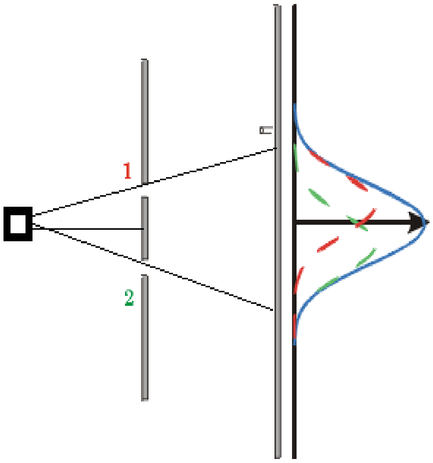


Figure 2.3: Double slit experiment

If the height of the two color curves represent the average number of particles arriving at each point on the screen, the red and green curves would represent, on average, the number of particles coming from the upper and lower slits, respectively. The sum of both functions (blue curve) would then be the total average number of particles hitting the screen at a given height. This curve shows us that a luminous band should be observed on the screen, whose highest intensity would be in the center, and would decrease as we move away from it. How then can the phenomenon of interference be explained in the corpuscular model?

One might think that each photon is able to pass through both slits simultaneously. However, by placing a small meter around one of the two slits, and if we produce photons individually and fire them one by one, we would observe that the detector would indeed be activated on average half of the time, each time a photon is detected passing through the corresponding slit. But the presence of the detector makes the interference phenomenon disappear immediately. It is as if nature behaved differently once it provides us with the information that would allow us, even in principle, to know the trajectory of each photon.

This interpretation seems far-fetched, although, as we shall see, it is the correct one. It would be more natural to think, as in fact it was done for decades, that the presence of the detector somehow disturbs the corresponding photon, and somehow interferes with it, which would end up destroying the fringe pattern observed before. Another more elementary explanation, very frequent in the literature, would be to suppose that light behaves as a wave, sometimes, and as a particle, other times, and this understanding occurs according to the circumstances.

But there is another experiment that shows that the disappearance of the interference fringes has nothing to do with the perturbation of the photons: let us place two linear polarizers: one horizontal, H, and the other vertical, V, just in front of the upper and lower slits, respectively.

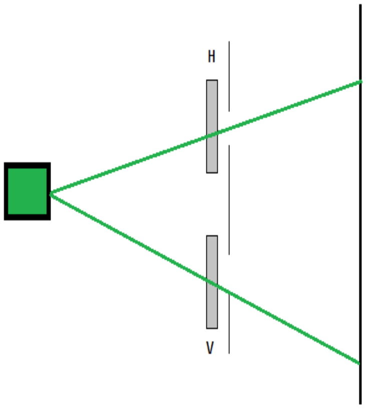


Figure 2.4: Double slit with polarizers

When observing the image formed on the screen, we can see that the fringe pattern disappears again. How can this phenomenon be explained? The reason is that the polarizers mark each photon: *H-type or V-type*. This would make it possible in principle to determine the path followed by the photon, because it would be enough to have placed at each point of the screen a polarizer of type H: If the photon passes through it, it is because it came from the upper slit; otherwise it would come from the lower slit. Access (even if only in principle) to this information is the reason for the disappearance of the interference pattern.

But it could still be argued that the polarizers interact with each photon by altering its properties. One way to show that this does not happen is to place a new polarizer at 45 degrees between the two slits and the screen:

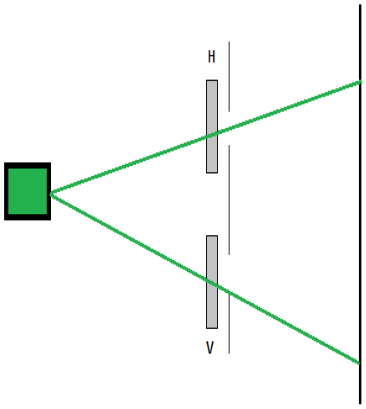


Figure 2.5: Double slit with three polarizers

In doing so, we observe that the interference pattern magically reappears. Note that this second polarizer could be placed at a sufficiently large distance so as not to cause any disturbance in the system, and yet the interference would still disappear. The interference appears again because the polarizer at 45 degrees erases the marking information making it impossible to determine the origin of the photon.

Let us see how the quantum formalism explains this phenomenon. First, and in the spirit of simplifying the discussion, let us assume that each photon is described by a vector $\phi = \sum_x \phi(x,t)e_x$ which at each instant t we must interpret as its amplitude function. The function ϕ is, of course, the solution to the Schrödinger wave equation that we will discuss in full later.

For the purpose of simplifying the discussion we have assumed that ϕ is a vector in a Hilbert space of finite dimension: each basis vector e_x is associated to a small vertical segment at height x such that $|\phi(x,t)|^2$ represents the probability that at instant t, and when making a measurement, the photon is at e_x, i.e., it is at a height between $x - \varepsilon$ and $x + \varepsilon$, measured from the midpoint between the two slits.

In doing so, we observe that the interference pattern magically reappears. Note that this second polarizer could be placed at a sufficiently large distance so as not to cause any disturbance in the system, and yet the interference would still disappear. The interference reappears because the polarizer at 45 degrees *erases the marking information*, *making it impossible to determine the origin of the photon.*

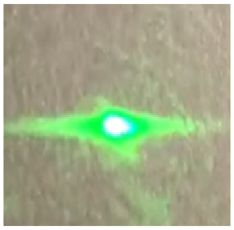


Figure 2.6: No interference pattern

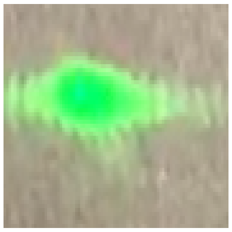


Figure 2.7: Interference pattern

After passing through the double slit we can write $\phi(x,t)$ as the sum of two vectors

$$\phi(x,t) = \phi_1(x,t) + \phi_2(x,t).$$

Intuitively this would correspond to the superposition of two (not necessarily orthogonal) states $\phi_1(x,t)$, which would mean "passing through the upper slit", and $\phi_2(x,t)$, which would be interpreted as "passing through the lower slit". Precisely, the wave functions ϕ_1 and ϕ_2 can be written as:

$$\phi_1(x,t) = \frac{1}{\sqrt{2\pi}\sqrt{(x+a/2)^2+(ct)^2}} e^{i\frac{h}{\lambda}\sqrt{(x+a/2)^2+(ct)^2}}$$
$$\phi_2(x,t) = \frac{1}{\sqrt{2\pi}\sqrt{(x-a/2)^2+(ct)^2}} e^{i\frac{h}{\lambda}\sqrt{(x-a/2)^2+(ct)^2}}$$

where a denotes the distance between the two slits, λ the photon frequency, h the Planck constant and x the vertical distance to the midpoint between the two slits.

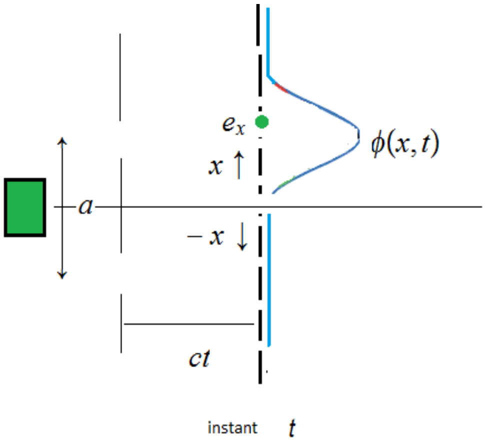


Figure 2.8: wave propagation

Thus, the probability of arriving at point x on the screen would be:

$$\begin{aligned}\left|\left\langle \sum_x (\phi_1(x,t)+\phi_2(x,t))e_x, e_x \right\rangle\right|^2 &= |\phi_1(x,t)+\phi_2(x,t)|^2 \\ &= (\phi_1(x,t)+\phi_2(x,t))(\overline{\phi_1}(x,t)+\overline{\phi_2}(x,t)) \\ &= |\phi_1(x,t)|^2+|\phi_2(x,t)|^2+\phi_1(x,t)\overline{\phi_2}(x,t)+\overline{\phi_1}(x,t)\phi_2(x,t) \\ &= |\phi_1(x,t)|^2+|\phi_2(x,t)|^2+\cos(\alpha_{12}(x,t)),\end{aligned}$$

where $\alpha_{12}(x,t)$ is the phase difference angle between ϕ_1 and ϕ_2 at point x, at instant t. It is this last term that explains the interference pattern.

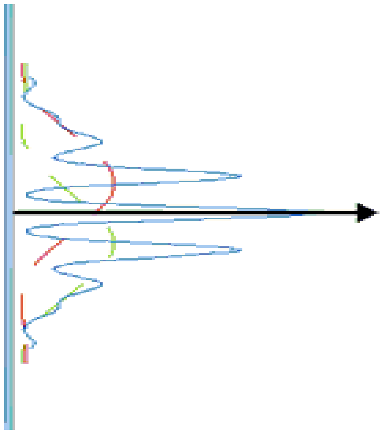


Figure 2.9: interference pattern

If photons behaved like classical corpuscles, the probability of reaching point x would be given by $|\phi_1(x,t)|^2 + |\phi_2(x,t)|^2$, which would be observed as a fringe that becomes less luminous as we move away from the center.

What happens now when two polarizers H and V are placed in front of each slit. The effect would be that each photon would be left in a mixed state that we could denote as a linear combination of basic states $e_x \otimes |H\rangle$ and $e_x \otimes |V\rangle$. After passing through the slits the mixed state of each photon would be expressible as

$$\phi(x,t) = \sum_x \phi_1(x,t) e_x \otimes |H\rangle + \sum_x \phi_2(x,t) e_x \otimes |V\rangle \,. \tag{2.5}$$

Note that in this sum there is no term of the form $\phi_1(x,t)e_x \otimes |V\rangle$ or of the form $\phi_2(x,t)e_x \otimes |H\rangle$ since no photon passing through the upper slit could be polarized vertically. Similarly, no photon passing through the lower slit could be polarized in the horizontal direction.

Now, the states $e_x \otimes |H\rangle$ y $e_{x'} \otimes |V\rangle$ are orthonogonal since

$$\langle e_x \otimes |H\rangle \,, e_{x'} \otimes |V\rangle\rangle = \langle e_x, e_{x'}\rangle \langle |H\rangle \,, |V\rangle\rangle = 0,$$

because $\langle |H\rangle \,, |V\rangle\rangle = 0$. From this we obtain that the following basis is orthonormal

$$B = \{e_x \otimes |H\rangle \,, e_{x'} \otimes |V\rangle\}$$

Then the probability that when collapsing (measuring) on the screen a given photon is at distance x from the center of the two slits would be:

$$|\phi_1(x,t)|^2 + |\phi_2(x,t)|^2 \,.$$

Consequently, the interference pattern disappears.

Let us denote by $|D_1\rangle$ and $|D_2\rangle$ the states corresponding to the polarizations in the 45 degree and $90 + 45$ degree directions. As we saw before:

$$|H\rangle = \frac{1}{\sqrt{2}} |D_1\rangle - \frac{1}{\sqrt{2}} |D_2\rangle$$
$$|V\rangle = \frac{1}{\sqrt{2}} |D_1\rangle + \frac{1}{\sqrt{2}} |D_2\rangle \,.$$

Then, the state (2.5) can be written in this basis in the following manner:

$$\begin{aligned}\phi(x,t) &= \sum_x \phi_1(x,t)e_x \otimes (\frac{1}{\sqrt{2}}\left|D_1\right\rangle - \frac{1}{\sqrt{2}}\left|D_2\right\rangle) + \sum_x \phi_2(x,t)e_x \otimes (\frac{1}{\sqrt{2}}\left|D_1\right\rangle + \frac{1}{\sqrt{2}}\left|D_2\right\rangle) \\ &= \frac{1}{\sqrt{2}}\sum_x(\phi_1(x,t)+\phi_2(x,t))e_x \otimes \left|D_1\right\rangle + \frac{1}{\sqrt{2}}\sum_x(\phi_2(x,t)-\phi_1(x,t))e_x \otimes \left|D_2\right\rangle .\end{aligned}$$

Measurement on the polarizer at 45 degrees placed between the slits and the screen is equivalent to collapsing the photon state $\phi(x,t)$ into one of two states.

$$\frac{1}{\sqrt{2}}\sum_x(\phi_1(x,t)+\phi_2(x,t))e_x, \tag{2.6}$$

$$\frac{1}{\sqrt{2}}\sum_x(\phi_2(x,t)-\phi_1(x,t))e_x. \tag{2.7}$$

In either case the probability of the photon reaching the screen at point x would be

$$\left|\phi_1(x,t)+\phi_2(x,t)\right|^2$$

or

$$\left|\phi_2(x,t)-\phi_1(x,t)\right|^2 ,$$

respectively, and, as we saw, the interference pattern must then be observed.

Chapter 3

Quantum Systems

3.1 The spin of a subatomic particle

3.2 Angular momentum

In order to understand the concept of spin of a particle in quantum mechanics we must first recall the classical notions of angular momentum and magnetic moment.

Suppose that a particle p with mass m moves along a simple closed trajectory in the x-y plane, parametrized by its radius vector $\overrightarrow{r}(t)$. We will denote by $r(t)$ its magnitude so that $\overrightarrow{r}(t) = r(t)e_r$, where e_r is the unit vector in the radial direction. We denote by $\phi(t)$ the angle p forms (counterclockwise) with respect to the x-axis.

Recall that its *angular momentum* is defined as the vector $\overrightarrow{j}(t) = \overrightarrow{r}(t) \times m\overrightarrow{v}(t)$, where $\overrightarrow{v}(t)$ denotes the velocity of the particle. The magnitude of this vector can be interpreted as the change in the function representing the area swept by p. In fact, the area swept by the particle between 0 and t is given by

$$A = \int_0^t \frac{1}{2} r(s)^2 \phi'(s) ds.$$

Then $A'(t) = \frac{1}{2}r^2(t)\phi'(t)$. On the other hand, in polar coordinates $\overrightarrow{j}$ may be written as

$$\begin{aligned} \overrightarrow{j}(t) &= r(t)e_r \times m\overrightarrow{v}(t) \qquad (3.1)\\ &= r(t)e_r \times m(r'(t)e_r + r(t)\phi'(t)e_\phi)\\ &= 2m\frac{r^2(t)}{2}\phi'(t)e_3 \simeq 2mA'(t)e_3, \end{aligned}$$

where e_ϕ is the unitary vector in the angular direction and $e_3 = e_r \times e_\phi$ is the standard unitary vector in the direction of the z axis.

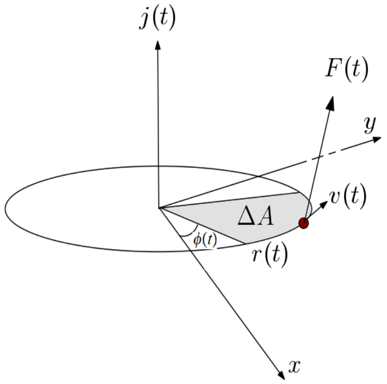


Figure 3.1: Angular momentum.

We recall that if p is moving under the action of a force $\overrightarrow{F}(t)$, its *torque* is defined as the vector

$$\overrightarrow{\tau}(t) = \overrightarrow{r}(t) \times \overrightarrow{F}(t).$$

This vector represents the change in angular momentum:

$$\overrightarrow{j}'(t) = \overrightarrow{r}(t) \times m\overrightarrow{v}'(t) + \overrightarrow{r}'(t) \times m\overrightarrow{v}(t) = \overrightarrow{r}(t) \times \overrightarrow{F}(t), \tag{3.2}$$

where the second term is zero since $\overrightarrow{r}'(t) = \overrightarrow{v}(t)$ and consequently their cross product is zero.

Let T denote the period of the particle, i.e., the time it takes p to make one complete turn. Suppose that the magnitude of the angular momentum is constant, say j. In this case we see from (3.1) that the total area enclosed by the path of p can be expressed as

$$A = \int_0^T A'(t)dt = \frac{1}{2m}\int_0^T jdt = \frac{T}{2m}j, \tag{3.3}$$

where e_ϕ is the unitary vector in the angular direction and $e_3 = e_r \times e_\phi$ is the standard unitary vector in the z-direction.

In the case $\overrightarrow{F}(t)$ is a central force (as gravity, for instance) the torque exerted by the force is zero and therefore $\overrightarrow{j}(t)$ is constant. Therefore, the area swept by the planet during an interval $\Delta t = t_2 - t_1$ would be

$$A = \int_{t_1}^{t_2} A'(t)dt = \frac{1}{2m}\int_{t_1}^{t_2} jdt = \frac{t_2 - t_1}{2m}j.$$

This is *Kepler's law*, that a planet in its orbit sweeps equal areas in the same period of time.

3.3 Magnetic momentum

Let us imagine a tiny rectangular-shaped wire ring (the shape is not relevant, but let us fix it for the purpose of making the discussion clearer) through which a stationary electric current flows, as shown in the following figure:

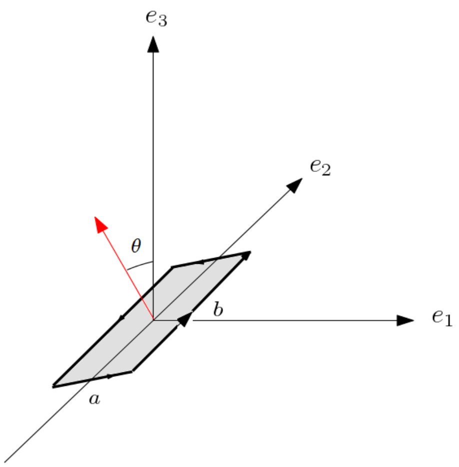


Figure 3.2: Magnetic momentum.

Suppose that present in the environment is a magnetic field $\overrightarrow{B}$ of constant magnitude B pointing in the direction of e_3. As current flows through the wire, each electron of charge q experiences a force due to $\overrightarrow{B}$ that is given by Lorentz's Law: $\overrightarrow{F} = q\overrightarrow{v} \times Be_3$, where q denotes the electron's charge and $\overrightarrow{v}$ its velocity. Let ΔQ denote the amount of charge contained in a small lateral segment of length Δl (see Figure 3.3). The total force experienced by this segment due to the field will be given by

$$\begin{aligned}\Delta\overrightarrow{F} &= \Delta Q\ \overrightarrow{v} \times \overrightarrow{B} = \rho A\Delta l\ v\ e_2 \times Be_3\\ &= v\rho A\ \Delta lBe_1,\end{aligned}$$

where ρ denotes the charge density present on the wire, A the area of its cross-section and v the magnitude of the vector $\overrightarrow{v}$.

By cutting the wire at any point, the amount of electric charge passing through the cross-section of area A in a time Δt can then be calculated as $v\rho\Delta tA$. *Current*, we recall, is defined as the amount of electric charge passing through the cross-section per unit time. Since $\Delta l = \Delta tv$ we can rewrite the equation above as $\Delta F = B\Delta lIe_1$. Consequently, the total force on the right lateral segment would then be

$$F = BbIe_1.$$

Similarly, over the left lateral segment the total force would be $-\overrightarrow{F}$.

Now, let us calculate the work W done by the force $\overrightarrow{F}$ on the right side of the loop to bring this side from position E to H:

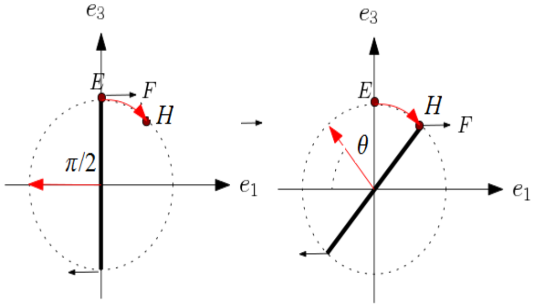


Figure 3.3

Notice that on the shorter sides of the loop the force is all directed in the direction normal to those two segments, and contained in the same plane as the loop, since $\overrightarrow{v} \times Be_3 = \pm e_2$, and consequently the total work on these two sides adds up to zero. On the other hand, on the right side of the loop the work would be given by the line integral

$$W_1 = \int_{\pi/2}^{\theta} \overrightarrow{F}(t) \cdot C'(t)dt,$$

where

$$\overrightarrow{C}(t) = \frac{a}{2}\cos(t)e_1 + \frac{a}{2}\sin(t)e_3$$

is the parametric equation of the circumference in the plane e_1-e_3 that traces the loop as it rotates. The calculation of W_1 is immediate:

$$\begin{aligned} W_1 = \int_{\pi/2}^{\theta} F \cdot C'(t)dt &= -BbI\frac{a}{2}\int_{\pi/2}^{\theta} \sin(t)dt \\ = \frac{Iab}{2}B\cos(t)\Big|_{\pi/2}^{\theta} &= \frac{Iab}{2}B\cos(\theta). \end{aligned}$$

Hence the total work is

$$W = 2W_1 = IabB\cos(\theta).$$

Let us note that the quantity $A = ab$ corresponds to the area enclosed by the circuit. It can be seen similarly that for an arbitrarily shaped loop the work done by B would be

$$W = IA\cos(\theta),$$

where A denotes the total area of the corresponding loop.

We define the *magnetic moment* of the loop as the vector $\overrightarrow{\mu} = (IA)n_\theta$. The work W can then be expressed as

$$W = \overrightarrow{\mu} \cdot \overrightarrow{B}.$$

Let us note that when $\overrightarrow{\mu}$ has the same direction as $\pm\overrightarrow{B}$, (this occurs if $\theta = 0$ or $\theta = \pi$) the value of W takes its minimum and maximum values, which correspond to the two possible equilibrium positions. On the other hand, when $\theta = \pi/2$, we see that W is zero. Hence the potential energy contained in the loop (in its original position) can be interpreted as $-W$, which corresponds to the work that would have to be exerted on the loop to overcome the torque of $\overrightarrow{F}$, that is, to bring it from the position where the normal vector forms an angle θ with B to that position where it forms an angle of $\pi/2$ (by definition the ground zero reference point):

$$U = -\overrightarrow{\mu} \cdot \overrightarrow{B}.$$

When work W is done against a force, the potential energy is defined as $U = -W$. The physical meaning of potential energy can be understood with a simple example. If for instance, work is done to lift a mass m a distance of h meters above a reference point, the amount of work done against gravity is equal to $W = -gmh$. Thus, the potential energy or accumulated energy of the mass would be $U = -W = gmh$ Newtons. If the particle lies h meters below the floor (point of reference) its potential energy would then be $-gmh$.

In the case of the loop we are considering, we take the vertical position of the loop as our reference point (that is why we measured the work done from $\theta = \pi/2$). The potential energy takes a minimal and a maximal value: A minimal value when $\theta = 0$, where $U = -IA$, and a maximal value when $\theta = \pi$, where $U = IA$.

Let us note, on the other hand, that the torque that $\overrightarrow{F}$ and $-\overrightarrow{F}$ exert on the lateral sides of the loop at any instant t can be easily calculated as $\overrightarrow{\mu}(t) \times \overrightarrow{B}$. In fact, for the right side of the loop, the torque exerted by $\overrightarrow{F}$ is equal to

$$\begin{aligned}\overrightarrow{\tau}_1(t) &= \left(\frac{a}{2}\cos(t)e_1 + \frac{a}{2}\sin(t)e_3\right) \times BbIe_1\\ &= \frac{abBI}{2}\sin(t)e_2.\end{aligned}$$

Similarly, the torque $\overrightarrow{\tau}_2(t)$ exerted by $-\overrightarrow{F}$ on the left vertical side is equal to $\overrightarrow{\tau_1}(t)$ and therefore the total torque would be

$$\overrightarrow{\tau}(t) = BabI\sin(\theta)e_2.$$

But clearly

$$\begin{aligned}\overrightarrow{\mu}(t) \times \overrightarrow{B} &= Iabn_\theta \times Be_3\\ &= Iab(-\sin(t)e_1 + \cos(t)e_3) \times Be_3\\ &= IabB\sin(t)e_2 = \overrightarrow{\tau}(t).\end{aligned} \tag{3.4}$$

3.4 Dipole formed by a particle

Let us suppose now that the dimensions of the loop are very small. Let us think that the whole system is reduced to a particle of mass m and charge q that rotates with enormous speed around a central point O tracing a circle of constant radius r and area A. This abstraction is called a *magnetic dipole*.

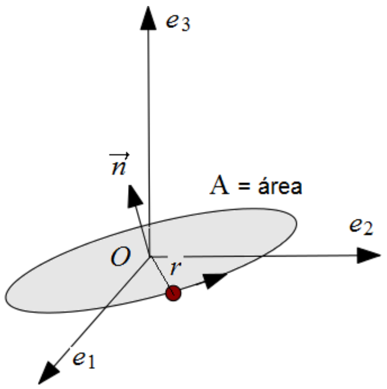


Figure 3.4: Magnetic dipole.

Its rotational period T is related to the current I flowing through the wire since q units of charge flow through each cross section of wire in T seconds. Hence, the current is equal to $I = q/T$ units of charge per second. Its angular momentum would then be

$$\overrightarrow{\mu}(t) = IA\overrightarrow{n}(t) = \frac{qA}{T}\overrightarrow{n}(t).$$

On the other hand, according to (3.3) its angular momentum $\overrightarrow{j}(t)$ would have magnitude

$$j = \frac{2mA}{T} = \frac{2mAI}{q} = \frac{2m}{q}\mu$$

where μ is the magnitude of the magnetic moment. This gives us the following *relationship between the magnetic moment and the angular momentum of the particle:*

$$\overrightarrow{\mu}(t) = \frac{q}{2m}\overrightarrow{j}(t). \tag{3.5}$$

The constant $\gamma = q/2m$ is called the *gyromagnetic ratio.*

We assume as before that our particle moves in the presence of a constant magnetic field $\overrightarrow{B} = Be_3$. Let $\theta(t) = \overrightarrow{j}(t) \cdot \overrightarrow{B}$ be the angle formed by its angular momentum with respect to $\overrightarrow{B}$. In (3.4) we calculated $\overrightarrow{\mu}(t) \times \overrightarrow{B}$ as the torque exerted by the force due to the magnetic field over the dipole. From equation (3.4) it follows that:

$$\overrightarrow{j}'(t) = \overrightarrow{\mu}(t) \times \overrightarrow{B} = \frac{qB}{2m}\overrightarrow{w}(t),$$

where $\overrightarrow{w}(t)$ is a vector perpendicular to $\overrightarrow{B}$ and to $\overrightarrow{\mu(t)}$. Since the potential energy of the dipole is constant, the magnitude of $\overrightarrow{w}(t)$ is also a constant, that we denote by w_0.

Now, since $\overrightarrow{j}'(t)$ is perpendicular to $\overrightarrow{B}$ one has

$$\frac{d}{dt}(\overrightarrow{j}(t) \cdot \overrightarrow{B}) = \overrightarrow{j}'(t) \cdot \overrightarrow{B} = 0.$$

Thus, the angle $\theta(t)$ is constant. Moreover, $\overrightarrow{j}(t)$ (and therefore $\overrightarrow{\mu}(t)$) moves around the z-axis at constant angular speed. This is because if $\phi(t)$ denotes the angle formed by $\overrightarrow{w}(t)$ between times t and $t + \Delta t$ one sees (see picture below) that

$$|\Delta j(t)| = |\overrightarrow{w}(t)|\,\phi(t) = w_0\phi(t)$$

and thus

$$\left|\overrightarrow{j}'(t)\right| = w_0\phi'(t)$$

This shows us that the magnetic momentum of the loop rotates in the circle described by $\overrightarrow{w}(t)$. with angular velocity given by

$$\phi'(t) = \frac{\left|\overrightarrow{j}'(t)\right|}{w_0} = \frac{\frac{qB}{2m}w_0}{w_0} = \frac{qB}{2m}. \tag{3.6}$$

This motion is called the *Larmour's precession of the magnetic moment.*

3.5 Magnetic moment of particles

Atoms possess magnetic moment, due to the rotation of their orbiting electrons. On the other hand, experiments show that subatomic particles such as electrons, protons, neutrons, etc., behave as if they were small magnetic dipoles, some kind of small rotating spheres spining about a certain

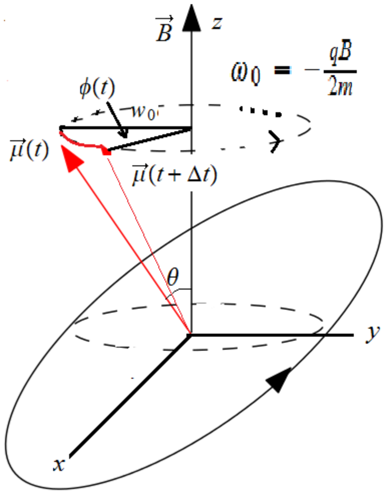


Figure 3.5: Larmour precession of an electron

axis. However, this analogy is useful only heuristically, because in fact some particles, even without electric charge, have magnetic moment.

Experimentally it can be seen that the gyromagnetic ratio of a particle is given by

$$\gamma = \frac{qg}{2m} \tag{3.7}$$

where g is a constant that depends on the the particle. For the electron, for example, the value of g is approximately 2; and for the proton $g \simeq 5.59$.

As we shall see, the experiment of Stern and Gerlach showed that the angular momentum of electrons has constant magnitude $j = \hbar/2$. Therefore, the magnetic moment has magnitude

$$\mu = \frac{\hbar}{2}\gamma.$$

In the presence of a magnetic constant field $\overrightarrow{B} = Be_3$ the potential energy associated to μ is then given by

$$\begin{aligned} U = -\mu \cdot B &= -\frac{\hbar}{2}\frac{qg}{2m} B \cos\theta \\ &= -\frac{\hbar}{2}\omega_0 \cos\theta, \end{aligned}$$

where ω_0 is, in analogy with (3.6), the angular velocity of the Larmor's precession:

$$\omega_0 = \frac{qg}{2m} B \tag{3.8}$$

3.6 Stern-Gerlach Experiment

Electrons as well as other subatomic particles behave as small dipoles. However, the values of the energy contained in them do not appear in a continuous way, *but only assume a set of discrete values.* This fact was demonstrated for the first time in a famous experiment carried out by the German physicists Otto Stern and Walther Gerlach, in 1922.

A furnace heats a cloud of silver atoms at a temperature of 1000 K which is then directed into a collimation slot of width 0.1mm. From there, a homogeneous beam of atoms exits at a speed of approximately 500 m/s. The beam is then passed through a Stern-Gerlach (SG) apparatus and finally each atom is made to collide against a screen located to the right of the device (see figure).

The apparatus consists of two powerful electromagnets capable of generating an *inhomogeneous* magnetic field varying only in the z-direction in a linear fashion. That is: $\partial B/\partial z = k \neq 0$, where the field strength grows in the direction of the z axis. The magnitude of B_z is very large (of the order of 10^4 Gauss) much larger than that of the horizontal components of the field, which we will assume to be negligible.

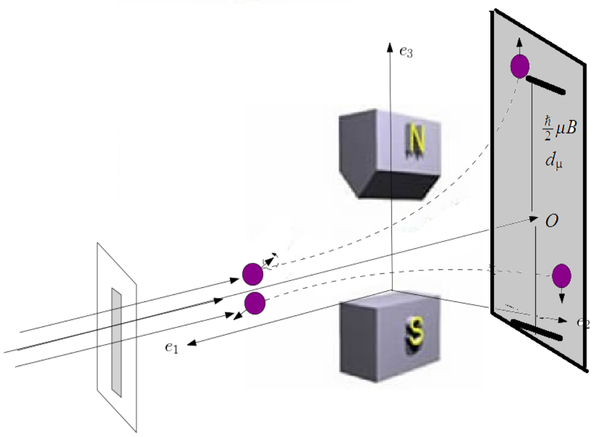


Figure 3.6: Stern-Gerlach.

Each silver atom contains 47 electrons, one of them isolated at the last orbital. This means that although it is electrically neutral (and therefore $\overrightarrow{B}$ does not exert any Lorentz force on each atom) its magnetic moment $\overrightarrow{\mu}$ is determined by this isolated electron. If $U = -\overrightarrow{\mu} \cdot \overrightarrow{B}$ is the potential energy of the particle, the force exerted on each atom is given by

$$\overrightarrow{F} = -grad(U) = grad(\overrightarrow{\mu} \cdot \overrightarrow{B}).$$

Then the magnitude of this force is given by:

$$F = k\frac{\hbar}{2}\omega_0 \cos\theta.$$

Under a constant force the trajectory of each atom depends on the value of $cos(\theta)$ and consequently each atom will follow either an upward or downward trajectory whose final destination will be a place of the screen located at a certain distance δ_θ from the central point O. Since the furnace prepares atoms in an isotropic manner, i.e., whose magnetic moments are randomly distributed, then we would expect to observe on the screen a series of impacts that would distribute continuously along the vertical direction, above and below O.

However, the experiment reveals only two impact stripes located at symmetric distances from the center. This result could be interpreted as the fact that atoms only come with magnetic moments in two possible configuration: up or down. In fact, it is observed that about 50 percent of the atoms come upwards and 50 downwards. A particle with this peculiarity is called a particle of *spin* 1/2.

To see how paradoxical this result is, let's place two devices of SG, one next to the other. The first one prepares spin 1/2 particles in the positive direction of the z-axis (the first device has a magnetic field in the direction of e_3). We then block those particles that reach the bottom of the screen and only let those of positive spin to continue their trajectory. Next, the particles that cross this first device are forced to pass through the second apparatus, whose magnetic field is tilted a small angle α with respect to the vertical. As we will see below, the probability that some particle passing through the second apparatus goes down is given by $\sin^2 \alpha$. If α is very close to zero this probability is very small, but not zero! Once in a while one could observe a particle moving downwards which is quite paradoxical since such particles have their magnetic moment directed in the positive vertical direction.

3.7 Spin 1/2

Consider again our SG apparatus with its magnetic field directed in the direction of e_3. The spin will then be (by definition) a quantum state represented by a complex unit vector of the form $\phi = \alpha\varepsilon_1 + \beta\varepsilon_2$, where ε_1 and ε_2 represent the *spin up*, and *spin down* states, respectively, when the spin is measured in the e_3 direction.

Having established this convention it remains for us to determine which two orthonormal vectors v_1, v_2 would correspond to the spin up and spin down states when the latter is measured in the direction of an arbitrary unit vector $\overrightarrow{n}$. That is, when using another SG apparatus whose magnetic field points in the direction of $\overrightarrow{n}$. It is clear that the basis change matrix between the orthonormal bases $\{\varepsilon_1, \varepsilon_2\}$ and $\{v_1, v_2\}$ should only depend on the relative position between the two SG apparatuses.

Suppose $\overrightarrow{B}$ is a magnetic field of constant magnitude B in the direction of the e_3 axis, and let a particle of spin 1/2 move through that field in a plane perpendicular to $\overrightarrow{B}$. Let $E_1 = \mu B$ and $E_2 = -\mu B$ be the two possible energy states of the magnetic dipole associated with the particle. We already know that

$$\gamma = \frac{qg}{2m}, \ \mu = \gamma\frac{\hbar}{2}, \ \omega_0 = \frac{qgB}{2m}.$$

Hence,

$$\begin{aligned} E_1 &= \mu B = \frac{\hbar}{2}\gamma B = \frac{\hbar}{2}\frac{qg}{2m}B = \frac{\hbar}{2}\omega_0 \\ E_2 &= -\frac{\hbar}{2}\omega_0 \end{aligned}$$

The energy levels E_1 and E_2 are known as *Zeeman's energy levels.* The Hamiltonian for the spin would be given by a diagonal matrix with entries E_1 and E_2. This matrix corresponds to a basis of eigenvector for these two eigenvalues.

$$H = \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} = \frac{\hbar}{2}\begin{bmatrix} \omega_0 & 0 \\ 0 & -\omega_0 \end{bmatrix}. \tag{3.9}$$

Let ε_1 and ε_2 be the corresponding eigenvectors of the associated Hamiltonian. The dynamic equation governing the particle spin $\phi(t) = \alpha(t)\varepsilon_1 + \beta(t)\varepsilon_2$ would then be:

$$\begin{bmatrix} \alpha'(t) \\ \beta'(t) \end{bmatrix} = \frac{-i}{\hbar}H\begin{bmatrix} \alpha(t) \\ \beta(t) \end{bmatrix} = \frac{-i}{2}\begin{bmatrix} \omega_0 & 0 \\ 0 & -\omega_0 \end{bmatrix}\begin{bmatrix} \alpha(t) \\ \beta(t) \end{bmatrix}.$$

Its solution is given by

$$\alpha(t) = \alpha_0 e^{\frac{-i\omega_0 t}{2}}, \ \beta(t) = \beta_0 e^{\frac{i\omega_0 t}{2}}, \tag{3.10}$$

where $\phi(0) = \alpha_0\varepsilon_1 + \beta_0\varepsilon_2$.

Let us now see what would be the Hamiltonian associated with the motion of the particle if the magnetic field $\overrightarrow{B}$ is of constant magnitude B but its direction is arbitrary: Let us write $\overrightarrow{B}$ in spherical coordinates:

$$\overrightarrow{B} = B\sin\theta\cos\phi e_1 + B\sin\theta\sin\phi e_2 + B\cos\theta e_3.$$

If we assume that the system is isolated, its Hamiltonian will be represented by a certain Hermitian matrix with constant coefficients.

$$H = \left[\begin{array}{cc} a & b \\ \overline{b} & c \end{array}\right].$$

with $a, c \in \mathbb{R}$.

The eigenvalues of H then correspond to the roots of the characteristic polynomial

$$\begin{aligned} p(t) = \det\left[\begin{array}{cc} a-t & b \\ \overline{b} & c-t \end{array}\right] &= (a-t)(c-t) - |b|^2 \\ = t^2 - (a+c)t + ac - \left|b^2\right| &= 0. \end{aligned}$$

It is clear that the two energy levels must be independent of the direction in which $\overrightarrow{B}$ points and therefore these must be equal, as in the previous case, to

$$E_1 = \mu B = \frac{\hbar}{2}\omega_0, \ E_2 = -\mu B = \frac{-\hbar}{2}\omega_0,$$

Therefore the sum of the two roots of $p(t)$, equal to $E_1 + E_2 = a + c$, must be equal to zero, hence it follows that $a = -c$. On the other hand, the product of the roots must be equal to $ac - |b^2|$ whence it follows:

$$\left(\frac{\hbar}{2}\omega_0\right)^2 = a^2 + |b|^2.$$

The most general solution of the above equation with real a is given by two complex numbers of the form $a = \frac{\hbar}{2}\omega_0\cos u(\theta)$, $b = \frac{\hbar}{2}\omega_0\sin u(\theta)e^{-v(\phi)i}$ for two functions u and v that depend continuously on θ and ϕ. Then H can be written as

$$H = \frac{\hbar\omega_0}{2}\left[\begin{array}{cc} \cos u(\theta) & e^{-v(\phi)i}\sin u(\theta) \\ e^{v(\phi)i}\sin u(\theta) & -\cos u(\theta) \end{array}\right]. \tag{3.11}$$

A direct calculation shows that the eigenvectors of H corresponding to $\hbar\omega_0/2$ and $-\hbar\omega_0/2$ are precisely:

$$\begin{aligned} v_1 &= e^{-iv(\phi)/2}\cos\left(\frac{u(\theta)}{2}\right)\varepsilon_1 + \sin\left(\frac{u(\theta)}{2}\right)e^{iv(\phi)/2}\varepsilon_2, \\ v_2 &= -e^{-iv(\phi)/2}\sin\left(\frac{u(\theta)}{2}\right)\varepsilon_1 + \cos\left(\frac{u(\theta)}{2}\right)e^{iv(\phi)/2}\varepsilon_2, \end{aligned} \tag{3.12}$$

respectively.

The change of basis matrix between $\{\varepsilon_1, \varepsilon_2\}$ and $\{v_1, v_2\}$ is given by

$$S(\theta,\phi) = \left[\begin{array}{cc} e^{iv(\phi)/2}\cos(u(\theta)/2) & -\sin(u(\theta)/2)e^{-iv(\phi)/2} \\ -\sin(u(\theta)/2)e^{iv(\phi)/2} & e^{-iv(\phi)/2}\cos(u(\theta)/2) \end{array} \right].$$

Since $S(\theta,\phi)$ comes from the Hamiltonian (3.11) it must satisfy that

$$S(\theta,\phi) = S(\theta,0)S(0,\phi).$$

This is because one can imagine that the apparatus A whose magnetic field is determined by the angles θ,ϕ is positioned in space by first rotating the original apparatus pointing in the direction of e_3 and angle ϕ around the z-axis, let's call it A_1 and then rotating A_1 an angle θ in the plane z-x' where x' is the x axis of A_1. Let us call this second apparatus A_2. Then one must imagine that inside each A_i the two beams are recombined using another pair of magnets so that *no measurement is performed.* The evolution of the spin state given by the Hamiltonian associated to A must then be the composition of what happens in A_1 followed by what happens inside A_2.

A similar argument shows that $S(\theta_1+\theta_2, 0) = S(\theta_2, 0)S(\theta_1, 0)$ and that $S(0,\phi_1+\phi_2) = S(0,\phi_2)S(0,\phi_1)$. Hence, in particular for any integer n one has $S(0, n\phi) = S^n(0,\phi)$ and $S(n\theta, 0) = S^n(\theta, 0)$. Notice that these two last equations imply that $u(n\theta) = nu(\theta)$ and that $v(n\phi) = nv(\phi)$. From this one also gets that

$$nu(\theta) = u(n\theta) = u(m\frac{n}{m}\theta) = mu(\frac{n}{m}\theta),$$

and therefore

$$u(\frac{n}{m}\theta) = \frac{n}{m}u(\theta).$$

Since $u(\theta)$ is a continuous function it must satisfy $u(r\theta) = ru(\theta)$ for any real number r. Thus, $u(\theta)$ is a linear function of θ.

In a similar fashion one deduces that $v(\phi)$ is also a linear function of ϕ. Since these two functions coincide at $\theta = 0$ (respectively at $\phi = 0$) with the identity maps one must have $u(\theta) = \theta$ and $v(\phi) = \phi$.

In terms of the components $B_1 = B\sin\theta\cos\phi$, $B_2 = B\sin\theta\sin\phi$, $B_3 = B\cos\theta$, the complex numbers a and b are written as

$$\begin{aligned} a &= \frac{\hbar}{2}\omega_0\cos\theta = -\mu B\cos\theta = -\mu B_3, && (3.13)\\ b &= \frac{\hbar}{2}\omega_0\sin\theta e^{-\phi i} = -\mu B \mathrm{sin}\theta e^{-\phi i} && (3.14)\\ &= -\mu B\sin\theta(\cos\phi - i\sin\phi) && (3.15)\\ &- \mu(B_1 - iB_2). && (3.16) \end{aligned}$$

Then the Hamiltonian matrix can be written as

$$H = -\mu(B_1\sigma_1 + B_2\sigma_2 + B_3\sigma_3), \tag{3.17}$$

where σ_i are the *Pauli matrices*

$$\sigma_1 = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right], \ \sigma_2 = \left[\begin{array}{cc} 0 & -i \\ i & 0 \end{array}\right], \ \sigma_3 = \left[\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array}\right].$$

Physicists often call this operator the *spin operator*

$$\boldsymbol{\mu} = \mu\boldsymbol{\sigma} = \frac{\hbar}{2}\gamma\ \boldsymbol{\sigma}, \text{ where } \boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3).$$

By this notation they refer to the fact that the Hamiltonian describing the spin measurement of a particle p with spin 1/2 in the direction of the field $\overrightarrow{B} = B_1e_1 + B_2e_2 + B_3e_3$ can be found as (3.17)

$$\begin{aligned} H &= -\boldsymbol{\mu}\cdot\overrightarrow{B} = -\frac{\hbar}{2}\gamma\ \ \boldsymbol{\sigma}\boldsymbol{\cdot}\, B \\ &= -\frac{\hbar}{2}\gamma(B_1\sigma_1 + B_2\sigma_2 + B_3\sigma_3). \end{aligned} \tag{3.18}$$

where $\boldsymbol{\sigma}\cdot B$ is understood as the matrix $B_1\sigma_1 + B_2\sigma_2 + B_3\sigma_3$.

3.8 Bloch Sphere

As we have just seen, the quantum state of a particle with spin 1/2 corresponds to a unit vector of $\mathbb{C}^2$. Let ε_1 and ε_2 be the standard basis vectors, and let $v = \alpha\varepsilon_1 + \beta\varepsilon_2$, with $|\alpha|^2 + |\beta|^2 = 1$. Let α and β be written in polar form $\alpha = r_1e^{\varphi_1 i}$, $\beta = r_2e^{\varphi_2 i}$. The fact that v is unitary tells us that $r_1^2 + r_2^2 = 1$. Since $0 \leq r_i \leq 1$, we can choose $\theta \in [0, \pi]$ such that $r_1 = \cos\left(\frac{\theta}{2}\right)$ y $r_2 = \sin(\frac{\theta}{2})$. We know that two vectors in $\mathbb{C}^2$ describe the same physical state if they differ by a non-zero complex multiple. Then the same quantum state can be represented as

$$\begin{aligned} v &= \cos\left(\frac{\theta}{2}\right)e^{\varphi_1 i}\varepsilon_1 + \sin\left(\frac{\theta}{2}\right)e^{\varphi_2 i}\varepsilon_2 \\ &= e^{\varphi_1 i}\left(\cos\left(\frac{\theta}{2}\right)\varepsilon_1 + \sin\left(\frac{\theta}{2}\right)e^{(\varphi_2-\varphi_1)i}\varepsilon_2\right) \\ &= \cos\left(\frac{\theta}{2}\right)\varepsilon_1 + \sin\left(\frac{\theta}{2}\right)e^{i\phi}\varepsilon_2, \end{aligned}$$

where $\phi = \varphi_2 - \varphi_2$. The state v is represented in the unit sphere S^2 in $\mathbb{R}^3$ (Bloch's sphere) as the vector with spherical coordinates (ϕ, θ)

The states of a two-dimensional quantum system are represented by elements of $\mathbb{CP}^1$, which is a space topologically identifiable with S^2. The Bloch sphere is an explicit manifestation of this fact. For example, the polarization state of a right-hand circularly polarized photon $1/\sqrt{2}\varepsilon_1 + i/\sqrt{2}\varepsilon_2$ corresponds to the point $\theta = \phi = \pi/2$. It is easy to see that a vector v and its orthogonal are represented by diametrically opposite points on the Bloch sphere.

Let us assume as before that $\overrightarrow{B} = Be_3$ is a constant magnetic filed in the direction of z. If the vector

$$v_1 = \cos\left(\frac{\theta_0}{2}\right)\varepsilon_1 + \sin\left(\frac{\theta_0}{2}\right)e^{i\phi_0}\varepsilon_2(3.12)$$

represents the initial state $v(0)$ of the spin of a particle moving in this field, at time t its state would be given by (3.10)

$$\begin{aligned} v(t) &= \cos\left(\frac{\theta_0}{2}\right) e^{\frac{-i\omega_0 t}{2}}\varepsilon_1 + \sin\left(\frac{\theta_0}{2}\right) e^{i\phi_0} e^{\frac{i\omega_0 t}{2}}\varepsilon_2 \\ &= \cos\left(\frac{\theta_0}{2}\right)\varepsilon_1 + \sin\left(\frac{\theta_0}{2}\right) e^{i(\phi_0+\omega_0 t)}\varepsilon_2, \end{aligned}$$

which corresponds to rotating $v(0)$ an angle ω_0 around the z-axis on the Bloch sphere. We see then that the spin has a precession motion as in the classical case.

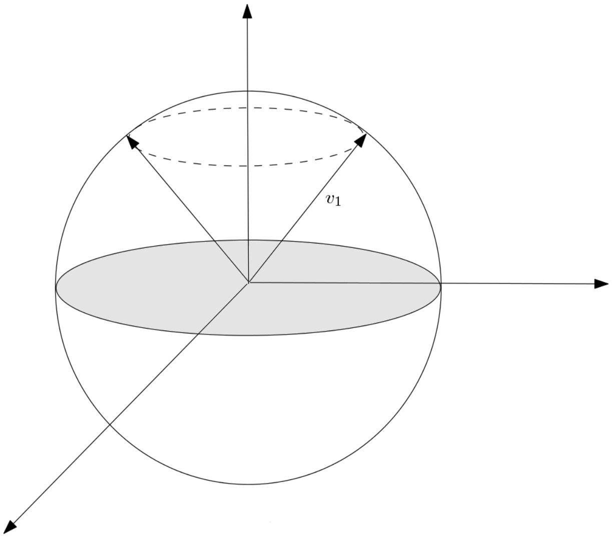


Figure 3.7: Spin precesion.

3.9 Rabi oscillations

Now we want to analyze how the spin moves in a magnetic field in which its horizontal component oscillates with a given frequency ω. Suppose that the magnetic field is given by $\overrightarrow{B} = b_0 e_3 + b_1(\cos(\omega t)e_1 - \sin(\omega t)e_2)$. According to (3.18) the Hamiltonian governing the spin state is given by

$$\begin{aligned} H &= -\frac{\hbar}{2}\gamma\,(B_1\sigma_1 + B_2\sigma_2 + B_3\sigma_3) \\ &= -\frac{\hbar}{2}\omega_0\sigma_3 - \frac{\hbar}{2}\omega_1(\cos(\omega t)\sigma_1 - \sin(\omega t)\sigma_2), \end{aligned}$$

where $\omega_0 = \gamma b_0$ y $\omega_1 = \gamma b_1$. In explicit form H can be written as

$$H = -\frac{\hbar}{2}\begin{bmatrix} \omega_0 & \omega_1 e^{i\omega t} \\ \omega_1 e^{-i\omega t} & -\omega_0 \end{bmatrix}. \tag{3.19}$$

In the case where ω is chosen equal to ω_0 the *magnetic resonance* phenomenon appears. Before proceeding to solve the Schrödinger equation with the Hamiltonian (3.19) let us anticipate the following fact from which the name resonance is derived: If the state of the spin at $t = 0$ is ε_1, the probability of finding it in the state ε_2 at an instant $t > 0$ is given by

$$p(t) = \left(\frac{\omega_1}{\Omega}\right)^2 \sin^2(\frac{\Omega t}{2}), \text{ con } \Omega = \sqrt{(\omega-\omega_0)^2+\omega_1^2}. \tag{3.20}$$

This is the so-called *Rabi oscillations* phenomenon. We see that the maximum amplitude is achieved when $\omega_1/\Omega = 1$, that is, if $\omega = \omega_0$, which is why this value is called the resonance frequency. As we will see below, for the resonance frequency the spin state $v(t)$ has the form.

$$v(t) = e^{i\omega_0 t/2}\ \widetilde{\alpha}(t)\varepsilon_1 + e^{-i\omega_0 t/2}\ \widetilde{\beta}(t)\varepsilon_2,$$

where

$$\begin{aligned}\widetilde{\alpha}(t) &= \alpha_0 \cos\left(\frac{\omega_1 t}{2}\right) + i\beta_0 \sin\left(\frac{\omega_1 t}{2}\right)\\ \widetilde{\beta}(t) &= i\alpha_0 \sin\left(\frac{\omega_1 t}{2}\right) + \beta_0 \cos\left(\frac{\omega_1 t}{2}\right),\end{aligned}$$

and where $v(0) = \alpha_0\varepsilon_1 + \beta_0\varepsilon_2$ is the original satate of the spin

3.9.1 Solution of the equation of motion

Let us proceed to solve the equation of motion for the spin of a particle moving under the action of the oscillating magnetic field defined by $\overrightarrow{B} = b_0 e_3 + b_1(\cos(\omega t)e_1 - \sin(\omega t)e_2)$. $\phi(t) = \alpha(t)\varepsilon_1 + \beta(t)\varepsilon_2$, Schrödinger's equation takes the form.

$$i\hbar \begin{bmatrix}\alpha'(t)\\ \beta'(t)\end{bmatrix} = -\frac{\hbar}{2}\begin{bmatrix}\omega_0 & \omega_1 e^{i\omega t}\\ \omega_1 e^{-i\omega t} & -\omega_0\end{bmatrix}\begin{bmatrix}\alpha(t)\\ \beta(t)\end{bmatrix}. \tag{3.21}$$

Define

$$\begin{aligned}\alpha(t) &= \widetilde{\alpha}(t)e^{i\omega_0 t/2}\\ \beta(t) &= \widetilde{\beta}(t)e^{-i\omega_0 t/2}\end{aligned}$$

Equations(3.21) can be written as

$$\begin{aligned}i\frac{d\widetilde{\alpha}(t)}{dt} &= -\frac{\omega_1}{2}e^{i(\omega-\omega_0)t}\widetilde{\beta}(t)\\ i\frac{d\widetilde{\beta}(t)}{dt} &= --\frac{\omega_1}{2}e^{-i(\omega-\omega_0)t}\widetilde{\alpha}(t).\end{aligned}$$

When $\omega = \omega_0$, (resonance condition) we see that from the latter system we can derive a second degree ordinary differential equation for $\widetilde{\alpha}(t)$

$$\frac{d^2\widetilde{\alpha}(t)}{dt^2} = -\frac{\omega_1^2}{4}\widetilde{\alpha}(t).$$

That is,

$$\frac{d^2\widetilde{\alpha}(t)}{dt^2} + \left(\frac{\omega_1}{2}\right)^2\widetilde{\alpha}(t) = 0.$$

From here we see that

$$\widetilde{\alpha}(t) = a\cos(\frac{\omega_1}{2}t) + b\sin(\frac{\omega_1}{2}t),$$

for two constants a, b that depend on the initial conditions. In a similar manner we see that

$$\widetilde{\beta}(t) = ia\sin(\frac{\omega_1}{2}t) - ib\cos(\frac{\omega_1}{2}t).$$

Therefore the solution sought is

$$v(t) = e^{i\omega_0 t/2}\ \widetilde{\alpha}(t)\varepsilon_1 + e^{-i\omega_0 t/2}\ \widetilde{\beta}(t)\varepsilon_2.$$

For $t = 0$ we see that $v(0) = a\varepsilon_1 - ib\varepsilon_2$. Hence, if $\phi(0) = \alpha_0\varepsilon_1 + \beta_0\varepsilon_2$ is the spin at its initial position, then

$$a = \alpha_0,\ \ b = \beta_0 i.$$

If we assume $v(0) = \varepsilon_1$ the spin evolution will be given by

$$v(t) = \cos\left(\frac{\omega_1 t}{2}\right) e^{i\omega_0 t/2}\varepsilon_1 + i\sin\left(\frac{\omega_1 t}{2}\right) e^{-i\omega_0 t/2}\varepsilon_2.$$

This last equation, except for the phase factor $e^{i\omega_0 t/2}$ can be written as

$$v(t) = \cos(\frac{\theta_t}{2})\varepsilon_1 + \sin(\frac{\theta_t}{2})e^{i\varphi_t}\varepsilon_2,$$

with

$$\omega_1 t = \theta_t,\ \ \varphi_t = -(\omega_0 t + \frac{\pi}{2}).$$

Hence, the probability of finding the spin in the ε_2 state is given by.

$$p(\varepsilon_1 \to \varepsilon_2) = \sin^2(\frac{\omega_1 t}{2}).$$

This is a function that oscillates between 0 and 1. These oscillations are known as Rabi oscillations, which as we saw correspond to the case $\Omega = \omega_1$ in (3.20).

On the other hand, since ω_1 and ω_0 can be controlled at will by varying the strength of the field $\overrightarrow{B}$ it becomes possible to take the spin to any point on the Bloch sphere and thus manipulate a given q-bit at our will. For example, if we want to take it from ε_1 to a state where the probability of finding it at ε_1 or ε_2 is equal, we can simply take $t_1 = \pi/(2\omega_1)$. At this instant the spin will be in the quantum state:

$$\begin{aligned} v(t) &= \frac{1}{\sqrt{2}}e^{i\omega_0 t_1/2}\varepsilon_1 + \frac{i}{\sqrt{2}}e^{-i\omega_0 t_1/2}\varepsilon_2 \\ &= \frac{1}{\sqrt{2}}\varepsilon_1 + \frac{i}{\sqrt{2}}e^{-i\omega_0 t_1}\varepsilon_2 \end{aligned}$$

So the probability of finding the particle's spin up or down is the same, when measured at time $t_1 = \pi/(2\omega_1)$.

Chapter 4

Quantum Computation

In this chapter we will give a brief introduction to quantum computation. As a main application we will discuss Shor's algorithm for factoring large integers.

4.1 Classical computation

Boolean functions and circuits

We start describing the classical model of computation, the technological implementation of a Turing machine.

A boolean circuit consists of a collection of wires carrying information in the form of a bit, a cero or a one, that pass trough a series of gates, which perform basic boolean computations. The classical truth tables can be interpreted abstractly as Boolean functions.

Definition 4.1.1. Let S denote the set consisting of the elements 0 and 1. A Boolean function of n arguments is a function $f : S^n \to S$ whose domain is the Cartesian product of n copies of S.

The truth tables of $\vee$, $\wedge$ and $\neg$, for example, can be viewed as Boolean functions, which we denote as follows:

x_1	x_2	$f(x_1, x_2) = x_1 \vee x_2$		x_1	x_2	$f(x_1, x_2) = x_1 \wedge x_2$
0	0	0		0	0	0
0	1	1	,	0	1	0
1	0	1		1	0	0
1	1	1		1	1	1

x	$f(x) = \neg x$
1	0
0	0

Similarly, the truth table of every formula $F(x_1, ..., x_n)$ is expressible as a function of propositional variables. It is clear that if F and G are equivalent formulas, i.e., if $F(x_{12}, ..., x_n) \Leftrightarrow G(x_1, ..., x_n)$, then F and G compute the same Boolean function.

The following proposition tells us that every Boolean function can be computed from the elementary functions $d(x_1, x_2) = x_1 \vee x_2$ and $n(x) = \neg x$.

In what follows, the disjunction (or conjunction) of functions h_i, with $i \in I$, will be denoted by $\vee h_i$ and by $\wedge h_i$, respectively.

Proposition 4.1.2. Let $f : S \times \cdots \times S \to S$ be an arbitrary boolean function. Then f can be written as a composition of functions $d(x_1, x_2) = x_1 \vee x_2$ and $n(x) = \neg x$.

Proof. First, we notice that $c(x_1, x_2) = x_1 \wedge x_2$ is clearly expressible as $c(x_1, x_2) = \neg(\neg x_1 \vee \neg x_2)$, which corresponds to the composition of functions $n(d(n(x_1), n(x_2)))$. By induction, one defines $c(x_1, \dots, x_n) = x_1 \wedge \cdots \wedge x_n$.

The domain of f can be divided into two different sets: U_0, the set of all tuples $(a) = (a_1, \dots, a_n)$ for which $f(a) = 0$; and U_1, the set of all tuples for which $f(a) = 1$. Fix $(a) \in U_1$. The characteristic function of (a) is defined as

$$g_{(a)}(x_1, \dots, x_n) = \varepsilon_1(x_1) \wedge \cdots \wedge \varepsilon_n(x_n),$$

where ε_i is the function $\varepsilon_i(x_i) = x_i$, if $a_i = 1$ and $\varepsilon_i(x_i) = \neg x_i$, if $a_i = 0$. (For instance, if $(a) = (1, 0, 0, 1)$ then

$$g_{(a)}(x_1, x_2, x_3, x_4) = x_1 \wedge \neg x_2 \wedge \neg x_3 \wedge x_4.$$

The key point is that $g_{(a)}$ is defined in such a way that $g_{(a)}(b) = 1$ if and only if $(b) = (a)$.

Let us see that

$$f(x_1, \dots, x_n) = \bigvee_{(a) \in U_1} g(a)(x_1, \dots, x_n).$$

We distinguish two cases:

1. If $(b) \in U_0$, then $g_a(b) = 0$, for all tuples $(a) \in U_1$ and henceforth

$$\bigvee_{(a) \in U_1} g_{(a)}(b) = 0 = f(b).$$

2. If $(b) \in U_1$, then $g_{(b)}(b) = 1$ and

$$\bigvee_{(a) \in U_1} g_{(a)}(b) = f(b) = 1.$$

The proposition thus follows immediately from (1) and (2). □

Example 4.1.3. Let f be the Boolean function defined by the following table:

x_1	x_2	$f(x_1, x_2)$	
0	0	0	
0	1	1	.
1	0	1	
1	1	0	

Let us express f as a composition of elementary functions. For each pair (a_1, a_2) with the property that $f(a_1, a_2) = 1$ we construct its corresponding characteristic function, according to the procedure explained in the proof of the previous proposition:

$$g_{(0,1)}(x_1, x_2) = \neg x_1 \wedge x_2$$
$$g_{(1,0)}(x_1, x_2) = x_1 \wedge \neg x_2.$$

Thus, $f(x_1, x_2) = (\neg x_1 \wedge x_2) \vee (x_1 \wedge \neg x_2)$.

The implementation of basic boolean function like $n(x)$, $d(x_1, \ldots, x_n)$ and $c(x_1, \ldots, x_n)$ are called *gates*, and are denote by

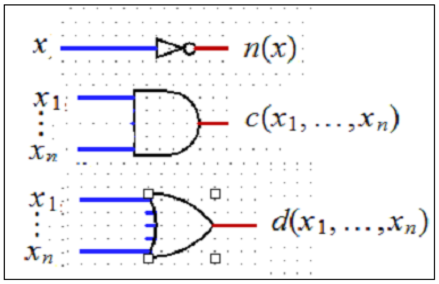
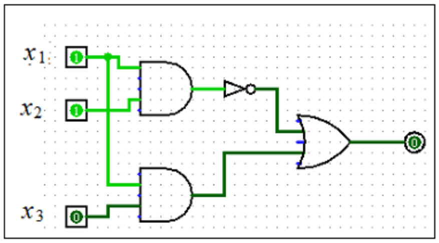


Figure 4.1: Boolean gates.

Gates are implemented by means of *electronic circuits* where a pulse of electricity is interpreted as 1, and its absence as 0.

For instance, the boolean functions $f(x_1, x_2, x_3) = \neg(x_1 \wedge x_2) \vee (x_1 \wedge x_3)$ can be implemented by the following circuit



$f(x_1,x_2,x_3) = x_1 \wedge x_2 \Rightarrow x_1 \wedge x_3$

Figure 4.2

For the entries $x_1 = 1$, $x_2 = 1$ (light green entries) and $x_3 = 0$ (dark green) the computation (at the right hand end) produces 0 (dark green). If you want to play with circuits you may use many free programs available in the web, like “Logisim” (http://logisim.uptodown.com/).

4.2 Quantum Computing

Quantum computation differs in many ways to classical computation. The analog of a bit will be the qubit, which instead of being just 0 or 1, it may be any unitary vector $v = \alpha e_0 + \beta e_1$ in the complex two dimensional Hilbert space $\mathbb{C}^2$. Similarly, n-bits are replaced by n-qubits, which correspond to unitary vectors in $H = \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$ (n-times). It is customary to use $|i_1, i_2, \ldots, i_n\rangle$ to denote the vector $e_{i_1} \otimes \cdots \otimes e_{i_n}$, where each i_k takes one of the values 0 or 1. This basis is called the standard or *computational basis* of H.

Quantum gates, on the other hand, are by definition unitary operators acting on qubits or quantum states. These operators can be represented by unitary matrices relative to the standard basis, and they are the building blocks of *quantum circuits*, a concatenation of unitary operators followed by a measurement at the end, once the last operator is applied. This last measurement collapses the superposition state into one of the vectors of the computational basis.

We recall that an unitary complex matrix U is a matrix such that $UU^* = I$, where U^* denote the conjugate transpose of U. Unitary matrices preserve the inner product, and therefore they preserve the norm hence transform n-qubits into n-qubits. Since unitary matrices are invertible, quantum gates are reversible, a property that classical gates do not posses in general.

Let us start by considering examples of 1-qubit gates.

1-qubit gates

A 1-qubit gate is a quantum gate which acts on unitary vectors of $\mathbb{C}^2$. These are the determined by their action on the computational basis $\{|0\rangle, |1\rangle\}$.

As an example, consider the operator, called the *NOT* gate, that sends $|0\rangle \mapsto |1\rangle$ and $|1\rangle \mapsto |0\rangle$. In the canonical basis this gate is represented by the unitary matrix

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

The quantum NOT gate is one of the four quantum 1-qubit gates known as *Pauli gates* that correspond to the Pauli matrices we introduced in (3.7).

$$\sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \ \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The linear operators represented by these matrices in the canonical basis $\{|0\rangle, |1\rangle\}$ are denoted by I, X, Y and Z, respectively. One can check right away that the last three correspond to rotations about x, y and z axes of the Bloch sphere.

It can be shown that the Pauli matrices generate the vector space of all 2×2 Hermitian matrices.

Another important example of a 1-qubit gate is the *Hadamard gate*. This is the unitary operator represented in the computational basis by the matrix

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

Suppose you have a single qubit $|\varphi\rangle \in \mathbb{C}^2$ and assume that you know that the quantum state of this qubit is either

$$|\varphi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$$

or

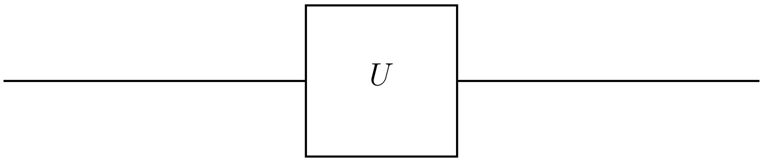
$$|\varphi\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle,$$

but you can not assure which of those two states it really is.
Making a measurement will destroy the state. However, by using the Hadamard gate we can discover with certainly the state of $|\varphi\rangle$. In fact, if $|\varphi\rangle = a|0\rangle + b|1\rangle$, where $a = 1/\sqrt{2}$ and $b = \pm 1/\sqrt{2}$, then

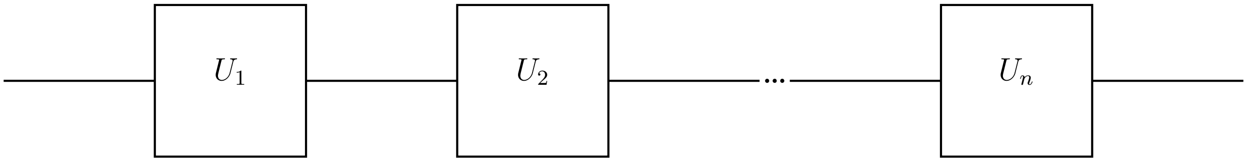
$$H|\varphi\rangle = \frac{1}{\sqrt{2}}(a+b)|0\rangle + \frac{1}{\sqrt{2}}(a-b)|1\rangle.$$

Hence, if we make a measurement after applying the Hadamard gate we get $|0\rangle$ with probability 1 when $|\varphi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$, and we obtain $|1\rangle$ with probability 1 when $|\varphi\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle$.

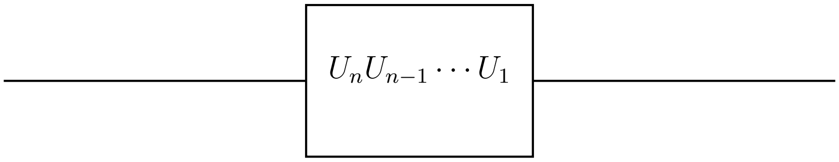
The 1-qubit gates are represented graphically by a box with an entry for the input and an exit for the output vector.



This means that given a single qubit $|\varphi\rangle$, the gate produces a new single qubit $U|\varphi\rangle$.
Since product of unitary matrices is a unitary matrix, the composition of quantum gates $U_1, \ldots, U_n$ is represented graphically by a sequence of boxes



This is of course equivalent to the quantum circuit



Multiple qubit gates

Now we are going to discuss the case of gates acting on n-qubits. In order to simplify the notation we will discuss the case of dimension $n = 2$.

Assume H_1 and H_2 are complex Hilbert spaces of dimension 2. The computational basis for $H_1 \otimes H_2$ is given by $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. Notice that if we identify $H_1 \otimes H_2$ with $\mathbb{C}^4$, the vector $|rs\rangle$ denotes the vector of $\mathbb{C}^4$ with one in the position $s \cdot 2^0 + r \cdot 2^1 + 1$ and zero otherwise. For example,

$$|10\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Given qubits or quantum states (unitary vectors) $|\varphi\rangle = a|0\rangle + b|1\rangle$ and $|\psi\rangle = c|0\rangle + d|1\rangle$, the state $|\varphi\rangle \otimes |\psi\rangle = |\varphi\rangle|\psi\rangle$, can be expressed as

$$|\varphi\rangle \otimes |\psi\rangle = |\varphi\rangle|\psi\rangle = ac|00\rangle + ad|01\rangle + bc|10\rangle + bd|11\rangle.$$

There are some quantum states in $H_1 \otimes H_2$ that can not be expressed as the product of two single qubits. For example, it is not possible to write the qubit $|\varphi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$ as the product of two single qubits $a|0\rangle + b|1\rangle$ and $c|0\rangle + d|1\rangle$. Suppose by contradiction that

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

then

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle = ac|00\rangle + ad|01\rangle + bc|10\rangle + bd|11\rangle$$

which implies that $ac = 1/\sqrt{2}$, $ad = 0$, $bc = 0$ and $bd = 1/\sqrt{2}$, thus $a = 0$ or $d = 0$. If $a = 0$ then $c \neq 0$ and therefore $b = 0$, which is a contradiction since $bd \neq 0$. This occurs, as we discussed before, when the two qubits are entangled.

Suppose that we have two 1-qubit gates, the NOT gate and the Hadamard gate given by matrices

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and

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

We can define a new quantum gate L given by the action of the NOT gate on the first 1-qubit and the Hadamard gate on the second 1-qubit. In order to find the unitary matrix that represents this gate we have to compute what L does on the elements of the computational basis of the tensor product space. For this, that $L|00\rangle = |01\rangle$, $L|01\rangle = |00\rangle$, $L|10\rangle = \frac{1}{\sqrt{2}}|10\rangle + \frac{1}{\sqrt{2}}|11\rangle$ and $L|11\rangle = \frac{1}{\sqrt{2}}|10\rangle - \frac{1}{\sqrt{2}}|11\rangle$. Therefore the matrix that represents this quantum gate in the canonical basis is

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

which is in fact an unitary matrix. This matrix is just the tensor product of the matrices σ_1 and H: $A = \sigma_1 \otimes H$. Therefore, we can compute the value of the quantum gate L applied on a quantum state $|\varphi\rangle = a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle$ just by multiplying the matrix A by the column vector whose entries are (from top to bottom) a, b, c, d.

On the other hand, we can also compute the action of the quantum gate L on $|\varphi\rangle$ as

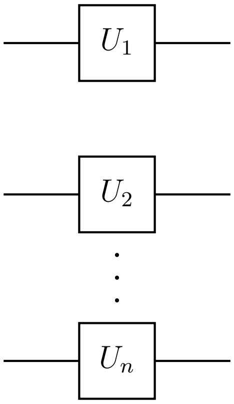
$$L|\varphi\rangle = a\sigma_1|0\rangle \otimes H|0\rangle + b\sigma_1|0\rangle \otimes H|1\rangle + c\sigma_1|1\rangle \otimes H|0\rangle + d\sigma_1|1\rangle \otimes H|1\rangle.$$

Remember that given two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ of size $m \times n$ and $k \times l$ respectively, the tensor product $A \otimes B$ is a matrix of size $mk \times nl$ defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}.$$

From the discussion above we see that given a multiple qubit in the space $H_1 \otimes \cdots \otimes H_n$, if we have 1-qubit gates $U_1, \ldots, U_n$ where U_i acts on the quantum states of H_i, then the tensor product $U_1 \otimes \cdots \otimes U_n$ defines a quantum gate acting on the multiple qubits of $H_1 \otimes \cdots \otimes H_n$.

Graphically, the quantum gate $U_1 \otimes \cdots \otimes U_n$ is represented as the quantum circuit



Controlled gates

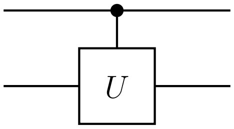
Another important class of gates in quantum computing are the controled gates. Let us start by discussing first controlled U-gates when U is a 1-qubit gate. Then we can see that it is not difficult to generalize this notion to arbitrary quantum circuits.

Suppose that we have a 1-qubit gate U. We define the controlled U-gate denoted by c-U as a 2-qubit gate acting in the following way: for $a, b \in \{0, 1\}$.

$$c - U|ab\rangle = \begin{cases} |a\rangle \otimes |b\rangle & \text{if} \quad a = 0 \\ |a\rangle \otimes U|b\rangle & \text{if} \quad a = 1 \end{cases}$$

.

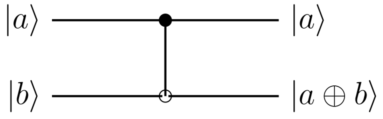
The c-U gates are represented in a quantum circuit with the following diagram:



Let us discuss the controlled c-NOT gate. We know that the NOT-gate is given in the canonical basis by the matrix $U = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Now the c-NOT gate is defined in the canonical basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ as $|00\rangle \mapsto |00\rangle$, $|01\rangle \mapsto |01\rangle$, $|10\rangle \mapsto |11\rangle$ and $|11\rangle \mapsto |10\rangle$. Therefore the c-NOT gate in this basis is represented by the 4×4 matrix

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

which is a unitary matrix since it is a *permutation matrix*.
The c-NOT gate is also represented by the quantum circuit



where $a, b \in \{0, 1\}$ and $a \oplus b$ denotes the sum $a + b$ modulo 2.

Measurements

In a quantum computer one is allowed to do any partial measurements of the qubits after a unitary operator is applied. For instance, if the domain of U is the Hilbert space $H_1 \otimes \cdots \otimes H_N$ one is allowed to measure some (or all) of the qubits at the right.

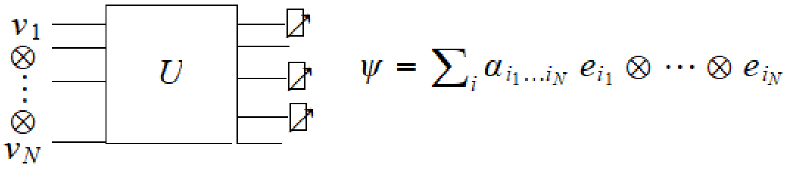

v_1
$\otimes$
$\otimes$
v_N
U
$\psi = \sum_i \alpha_{i_1 \ldots i_N} e_{i_1} \otimes \cdots \otimes e_{i_N}$

Figure 4.3: Measurement.

We may imagine that any quantum computation is carried out as follows: We prepare n particles of spin 1/2, each one in some state $v_i \in \mathbb{C}^2$. The box representing the operator U will be some isolated cavity where electromagnetic pulses transform the original quatum state of the system $\phi = v_1 \otimes \cdots \otimes v_N$ into some other state

$$\psi = \sum_i \alpha_{i_1 \ldots i_N} e_{i_1} \otimes \cdots \otimes e_{i_N}.$$

We then measure the corresponding spins, all at the same time, or one by one at each step at a time in any order. The final outcome must be the same. For instance, assume $N = 2$ and suppose that the final state after applying U is given by the unitary vector

$$\psi = \alpha_{00} e_0 \otimes e_0 + \alpha_{01} e_0 \otimes e_1 + \alpha_{10} e_1 \otimes e_0 + \alpha_{11} e_1 \otimes e_1$$

If we measure the spin of the first particle, and let's say the outcome is e_0, then the state of the system will collapse into

$$\psi_1 = \frac{\alpha_{00}}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}} e_0 \otimes e_0 + \frac{\alpha_{01}}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}} e_0 \otimes e_1$$

with probability $p_1 = |\alpha_{00}|^2 + |\alpha_{01}|^2$. If we then measure the spin of the second particle, and if we obtain for instance e_1, the state ψ_1 will collapse into $\psi_2 = e_0 \otimes e_1$, with probability

$$p_2 = \frac{|\alpha_{01}|^2}{|\alpha_{00}|^2 + |\alpha_{01}|^2}.$$

Thus, we see that the probability of collapsing into ψ_2 will be $p_1 \times p_2 = |\alpha_{01}|^2$. This is the same probability we would have obtained if we had measured both spins at the same time.

As more specific example, in the EPR experiment (2.4), the original state of the two photons is given by the vector

$$\begin{aligned} \eta = {} & \frac{1}{\sqrt{2}} \cos(\alpha - \beta) e_1(\alpha) \otimes e_1(\beta) + \frac{1}{\sqrt{2}} \cos(\alpha - \beta) e_2(\alpha) \otimes e_2(\beta) + \\ & + \frac{1}{\sqrt{2}} \sin(\alpha - \beta) e_1(\alpha) \otimes e_2(\beta) - \frac{1}{\sqrt{2}} \sin(\alpha - \beta) e_2(\alpha) \otimes e_1(\beta)). \end{aligned} \tag{4.1}$$

But after Alex measures the passing of his photon, the state partially collapses (with probability 1/2) into

$$\eta_1 = \cos(\alpha-\beta)e_1(\alpha)\otimes e_1(\beta) + \sin(\alpha-\beta)e_1(\alpha)\otimes e_2(\beta).$$

Then Beatrice measures the polarization of her photon and obtains $e_1(\beta)$ (it goes through the filter) with probability $\cos^2(\alpha-\beta)$. Thus, the probability of both events is $1/2\cos^2(\alpha-\beta)$, the same as the probability of collapsing into $e_1(\alpha)\otimes e_1(\beta)$ after one single measurement of the polarization of both photons when performed at the same time.

4.3 Computing boolean functions with a quantum computer

As before, we denote by S the set consisting of the two bits 0 and 1. Suppose that $f : S^n \to S^m$ is any boolean function. As we noticed before, f cannot be computed directly by means of a unitary transformation since in general f is not injective, hence not reversible as a computation. However, there is a trick that allows us to compute it quantum mechanically by introducing auxiliary qubits.

Before we do this, let us introduce a notation that is very useful in what follows. By H we denote the Hilbert space $\mathbb{C}^2$ and by $H^{\otimes n}$ the tensor product $H\otimes\cdots\otimes H$ of n copies of H. The standard basis for $H^{\otimes n}$ is formed by the vectors $e_{(i)} = e_{i_0}\otimes\cdots\otimes e_{i_{n-1}}$, with $i_\nu = 0, 1$. There are of course 2^n elements that we may numerated by using binary numbers from 0 to $2^n - 1$. Hence, the vector $e_{(i)}$ can be denoted by $|i_{n-1}\dots i_1 i_0\rangle$ or more simply by $|i\rangle$. If $n = 3$, for instance, the computational basis of $H^{\otimes 3}$ will be denoted by

$$\{|000\rangle\,,\ |001\rangle\,,\ |010\rangle\,,\ |011\rangle\,,\ |100\rangle\,,\ |101\rangle\,,\ |110\rangle\,,\ |111\rangle\}.$$

Elements of $H^{\otimes n}\otimes H^{\otimes m}$ will also be denoted by $|i\rangle\otimes|j\rangle$.

Now we explain how to compute f as above. For this, we define $B_f : H^{\otimes n}\otimes H^{\otimes m}\to H^{\otimes n}\otimes H^{\otimes m}$ the linear operator defined on the computational basis as:

$$B_f(|i\rangle\otimes|j\rangle) = |i\rangle\otimes|j\oplus f(i)\rangle$$

This operator will be represented as

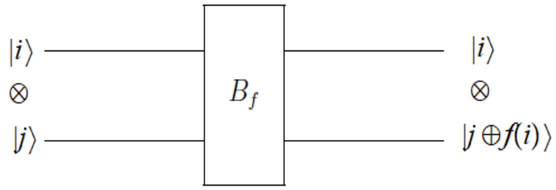


Figure 4.4: Computing a classical function

in Dirac's notation.

The operator B_f permutes the standard basis of $H^{\otimes n}\otimes H^{\otimes m}$. In fact, if

$$B_f(|i\rangle\otimes|j\rangle) = B_f(|r\rangle\otimes|s\rangle)$$

then $i = r$ and also $j\oplus f(i) = s\oplus f(r)$. By adding $f(r) = f(i)$ on both sides we obtain $j = s$.

Since any permutation operator is clearly unitary we see that B_f is unitary.

4.4 The Deutsch-Josza algorithm

The Deustch-Joza algorithm was one of the first algorithms implemented in a quantum computer. The problem, one has to admit, is rather artificial, designed specifically to illustrate the advantages of a quantum computer over a classical computer.

Let $f : S^n \to S$ be a boolean function. Suppose we know in advanced that f is of one of the two following types. Either f is constant or f is *balanced*, which means that for half the values of S^n the function takes the value zero, and for the other half the value one. That is

$$|\{x \in S^n : f(x) = 1\}| = |\{x \in S^n : f(x) = 0\}| = \frac{1}{2^{n-1}}.$$

We assume there is some sort of device or black box that gives $f(x)$ for each input $(x) \in S^n$. Equivalently, we may assume that we are given a gate of the form

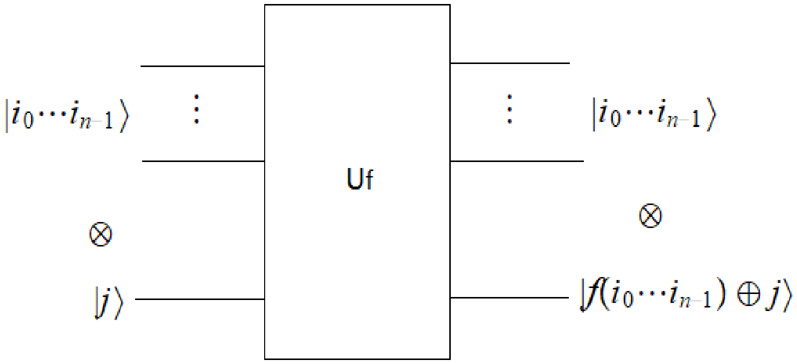

$|i_0 \cdots i_{n-1}\rangle$
Uf
$|i_0 \cdots i_{n-1}\rangle$
$\otimes$
$\otimes$
$|j\rangle$
$|f(i_0 \cdots i_{n-1}) \oplus j\rangle$

Figure 4.5

Our problem is to determine if f is constant or if f is balanced.

To solve this problem classically we would have to use our black box 2^{n-1} times, in the worst of cases. In a quantum computer, however, let us see that just one time is enough. For this, we use the following architecture of gates

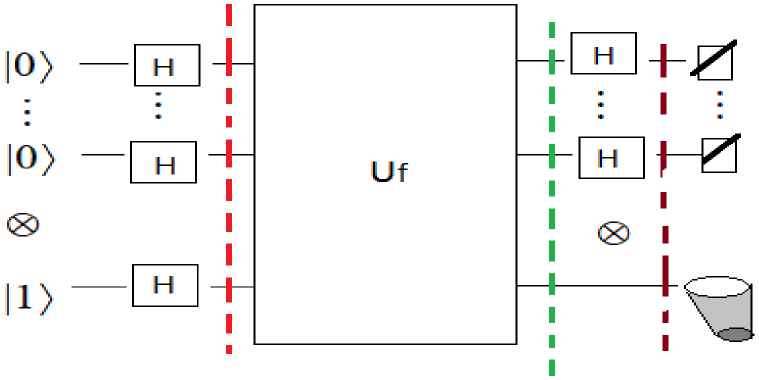


Figure 4.6

Each one of the little boxes is a Hadamard operator. On the right, at the output, we perform a measurement on the first n-qubits and disregard the last one.

In order to understand how this set of gates work, we first make some remarks on the tensor product of hadamard operators.

For a 1-qubit operator H we know that

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

This can be written succinctly as

$$H\left|x\right\rangle = \frac{1}{\sqrt{2}}\left|0\right\rangle + (-1)^x\frac{1}{\sqrt{2}}\left|1\right\rangle = \frac{1}{\sqrt{2}}\sum_{y\in\{0,1\}}(-1)^{xy}\left|y\right\rangle .$$

For a tensor product of n Hadamard operators, $H^{\otimes n}$ one has a similar formula, as one can easily check.

$$H\left|x_{n-1}\cdots x_0\right\rangle = \frac{1}{\sqrt{2^n}}\sum_{y_\nu}\in\{0,1\}(-1)^{x_0y_0+\cdots+x_{n-1}y_{n-1}}\left|y_{n-1}\cdots y_0\right\rangle .$$

Or, in compact form

$$H\left|x\right\rangle = \frac{1}{\sqrt{2^n}}\sum_{y\in\{0,1\}^n}(-1)^{x\cdot y}\left|y\right\rangle .$$

Let us analyze step by step how the input $\phi_0 = \left|0\cdots 0\right\rangle\otimes\left|1\right\rangle$ is transformed at the three different levels (red, green, brown). After the first set of Hadamard gates this input is transformed into

$$\begin{aligned} \phi_1 &= (H^{\otimes n}\otimes H)\left|0\cdots 0\right\rangle\otimes\left|1\right\rangle = \frac{1}{\sqrt{2^n}}\sum_{z\in\{0,1\}^n}\left|z\right\rangle\otimes\left(\frac{1}{\sqrt{2}}\left|0\right\rangle - \frac{1}{\sqrt{2}}\left|1\right\rangle\right) \\ &= \frac{1}{\sqrt{2^n}}\sum_{z\in\{0,1\}^n}\frac{1}{\sqrt{2}}\left|z\right\rangle\otimes\left|0\right\rangle - \frac{1}{\sqrt{2^n}}\sum_{z\in\{0,1\}^n}\frac{1}{\sqrt{2}}\left|z\right\rangle\otimes\left|1\right\rangle . \end{aligned}$$

After passing through the box U_f (green line) we obtain

$$\phi_2 = (H^{\otimes n}\otimes H)\phi_1 = \frac{1}{\sqrt{2^n}}\sum_{z\in\{0,1\}^n}\frac{1}{\sqrt{2}}\left|z\right\rangle\otimes\left|0\oplus f(z)\right\rangle - \frac{1}{\sqrt{2^n}}\sum_{z\in\{0,1\}^n}\frac{1}{\sqrt{2}}\left|z\right\rangle\otimes\left|1\oplus f(z)\right\rangle .$$

Each sum can be separated into two sums, one for those $z\in O$ such that $f(z)=0$ and those $z\in I$ for which $f(z)=1$

$$\begin{aligned} \phi_2 &= \frac{1}{\sqrt{2^n}}\sum_{z\in O}\frac{1}{\sqrt{2}}\left|z\right\rangle\otimes\left|0\right\rangle + \frac{1}{\sqrt{2^n}}\sum_{z\in I}\frac{1}{\sqrt{2}}\left|z\right\rangle\otimes\left|1\right\rangle \\ &\quad -\frac{1}{\sqrt{2^n}}\sum_{z\in O}\frac{1}{\sqrt{2}}\left|z\right\rangle\otimes\left|1\right\rangle - \frac{1}{\sqrt{2^n}}\sum_{z\in I}\frac{1}{\sqrt{2}}\left|z\right\rangle\otimes\left|0\right\rangle . \end{aligned}$$

In compact form this expression can be written as

$$\begin{aligned} \phi_2 &= \frac{1}{\sqrt{2^n}}\sum_{z\in\{0,1\}^n}\frac{(-1)^{f(z)}}{\sqrt{2}}\left|z\right\rangle\otimes\left|0\right\rangle + \frac{1}{\sqrt{2^n}}\sum_{z\in\{0,1\}^n}\frac{(-1)^{f(z)}}{\sqrt{2}}\left|z\right\rangle\otimes\left|1\right\rangle \\ &= \frac{1}{\sqrt{2^n}}\sum_{z\in\{0,1\}^n}(-1)^{f(z)}\left|z\right\rangle\otimes\left(\frac{1}{\sqrt{2}}\left|0\right\rangle - \frac{1}{\sqrt{2}}\left|1\right\rangle\right) . \end{aligned}$$

Finally, at the level of the brown line one has

$$\phi_3 = (H^{\otimes n} \otimes Id)(\phi_2) = \frac{1}{\sqrt{2^n}} \sum_{\substack{z\in\{0,1\}^n \\ w\in\{0,1\}^n}} (-1)^{f(z)+z\cdot w} |w\rangle \otimes \left(\frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle \right).$$

After *disregarding* the last qubit one gets

$$\phi_4 = \sum_{w\in\{0,1\}^n} \frac{1}{2^n} \left(\sum_{z\in\{0,1\}^n} (-1)^{f(z)+z\cdot w} \right) |w\rangle .$$

Hence, after measuring the first n qubits the probability that ϕ_4 collapses on $|w\rangle$ is given by

$$p = \left| \frac{1}{2^n} \sum_{z\in\{0,1\}^n} (-1)^{f(z)+z\cdot w} \right|^2 .$$

When $|w\rangle = |0\cdots 0\rangle$, this probability p equals 1, if f is constant, and 0 if it is balanced. Hence, if f is constant we must measure $|0\cdots 0\rangle$ with certainty. Thus, if f were balanced one would obtain as a result of the measurement a value $|w\rangle \neq |0\cdots 0\rangle$.

Remark 4.4.1. "Disregarding" a qubit, in a precise manner, means the following. If the corresponding quantum state

$$\phi = \sum \alpha_{(i)} e_{i_0} \otimes \cdots \otimes e_{i_{n-1}} = v \otimes e_0 + v \otimes e_1,$$

with

$$v = \sum_{(j)} \alpha_{(j)} e_{i_0} \otimes \cdots \otimes e_{n-2},$$

represents, for instance, the state of n particles of spin 1/2, we measure the spin of the n-th particle, and ignore the value we find. We only care about the state of the other $n-1$ particles, that collapses into $\frac{\sqrt{2}}{2}v$.

4.5 Shor's algorithm

In this section we want to explain what is arguably the most celebrated algorithm in quantum computation. Its main goal is to factor a very large integer, a computation nobody knows yet if could be efficiently implemented in a classical computer. The details of this algorithm can be read in any of the standard references, for instance [Chung]. We will rather concentrate on explaining the main ideas behind it.

The core of Shor's procedure relies on the possibility of computing the order of an element in the multiplicative group of those integers relatively prime to N, a group we will denote by $\mathbb{Z}_N^*$. Factoring an integer reduces to computing the order of an element a in this group, that is, to computing the least exponent r such that $a^r = 1 \bmod N$ (see [Chung]).

The main tool is the Quantum Fourier Transform that, except for a factor, it is the same as the classical discrete Fourier transform DFT.

The Fourier Transform

Let G be any finite abelian group of orden n. The space of all complex valued functions on G, that we denote by $L^2(G)$, can be given a structure of a Hilbert space if we define an inner product by

$$\langle f, g\rangle = \frac{1}{n}\sum_{x\in G}\overline{f(x)}g(x).$$

With this product, $L^2(G)$ is a complex Hilbert space of dimension n. There are two natural orthonormal bases for this space. One is the basis of delta functions $\delta_a(x)$, $a \in G$, where $\delta_a(x) = 1$ only at the value $x = a$. The other one is the *Fourier* basis which consists of the characters of the group. In this section we will only deal with the case of the cyclic group $G = \mathbb{Z}_N$ representing a circle or a regular polygon with n vertices. For this group, it can be seen ([TA], Chapter 2) that the characters are given by the exponential functions

$$e_a(x) = e^{\frac{2\pi i}{N}xa}, \quad a \in G.$$

It is easy to show that this is indeed an orthonormal basis for $L^2(G)$. In this basis any function f can be written as

$$f(x) = \sum_{a\in G}\langle f, e_a\rangle\, e_a(x).$$

The function that computes the coefficients $\langle f, e_a\rangle$ is called the *Discrete Fourier Transform* of f, (DFT) and it is denoted by $\widehat{f}$. More explicitly

$$\widehat{f}(a) = \langle f, e_a\rangle \text{, for each } a \in G.$$

The study of the Fourier transform is a whole branch of mathematics. However, for our modest purposes we just need to know that the DFT is notable because of its ability to detect cyclic patters.

Suppose that r divides N. Let f be the function on G such that $f(x) = 1$ only if x is a multiple of $r : x = 0, r, 2r, 3r, \ldots, (\frac{N}{r} - 1)r$. The function f encodes a binary sequence with a 1 repeating with period r, and zeroes elsewhere. Let us compute its Fourier transform

$$\begin{aligned}\widehat{f}(a) &= \langle f, e_a\rangle = \frac{1}{N}\sum_{x\in G}\overline{f(x)}e^{\frac{2\pi i}{N}xa} = \frac{1}{N}\sum_{v=0}^{N/r-1} e^{\frac{2\pi i}{N}vra} \\ &= \frac{1}{N}\sum_{v=0}^{N/r-1}(\omega^{ra})^v,\end{aligned}$$

where $\omega = e^{\frac{2\pi i}{N}}$ is a primitive Nth-root of unity. If a is not a multiple of r/a, the last sum can be computed as

$$\widehat{f}(a) = \frac{1}{N}\frac{\omega^{aN} - 1}{\omega^{ra} - 1}.$$

We see that if $a = vN/r$ is a multiple of N/r then the value of this sum is equal to $1/r$. Otherwise, it is zero. What this means is that the DFT of f captures the cyclic pattern of f by detecting a peak of "high frequency" at those points that are multiples of N/r.

The quantum Fourier transform QFT, on the other hand, is defined as above but changing the normalization factor by $1/\sqrt{N}$. Formally, if $\{x_1, \ldots, x_N\}$ is a sequence of complex numbers, the QFT of this sequence is another sequence $\{y_1, \ldots, y_N\}$ where

$$y_a = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1}\overline{x_k}e^{2\pi i k a/N}. \tag{4.2}$$

Quantum circuit

Now let us describe a quantum circuit that computes the order of an element a in $\mathbb{Z}_N^*$. Shor's quantum circuit looks like this:

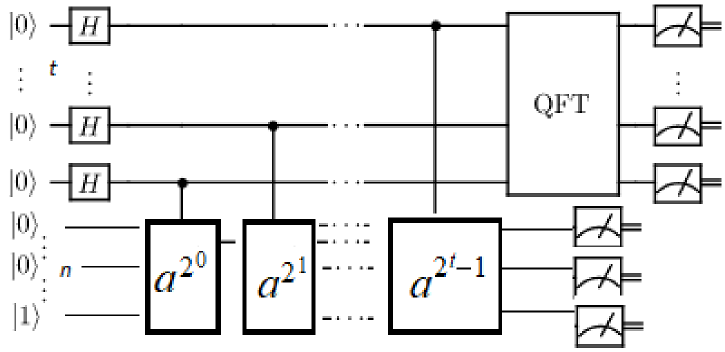


Figure 4.7: Shor's circuit

It contains two registers. The top register has t-qubits at its entrance while the bottom register has n-qubits. The integer n is taken to be the smallest power of 2 greater or equal than N, that is $n = \lfloor \log_2(N) \rfloor + 1$. The integer t, on the other hand, is an auxiliary integer that serves as a calibration parameter. It is roughly taken to be of size $t = 2n + 1$. The entrance of both registers are interpreted as binary numbers, form 0 to $2^t - 1$, for the first register, and from 0 to $2^n - 1$ for the second one.

Each one of the boxes labeled with an H represents a Hadamard gate. The boxes in the second register represent controlled multiplication by a, a^2, …,a^{2^t-1}, in that order. Hence, if the zeroth qubit of the top register (counted from bottom to top) is equal to $j_0 = 1$ we perform the operation *multiplication by a*. If, on the other hand, $j_0 = 0$, we do nothing (apply the identity operation). Similarly, the box labeled by a^2 is activated only if the first qubit $j_1 = 1$. We proceed in a similar fashion for the other multiplication boxes.

Finally, the last box on the right represents the computation of the QFT applied to the qubits of the first register.

Let us follow the circuit step by step. After the first t Hadamard gates the state of the system will be

$$\phi = \frac{1}{\sqrt{2^t}} \sum_{j=0}^{2^t-1} |j_{t-1}, \ldots, j_0\rangle \otimes |0 \ldots 01\rangle .$$

We will write this state simply as

$$\phi = \frac{1}{\sqrt{2^t}} \sum_{j=0}^{2^t-1} |j\rangle \otimes |1\rangle .$$

Then, after we encounter the box *multiplication by* a^{2^0} the state becomes

$$\phi_0 = \frac{1}{\sqrt{2^t}} \sum_{j=0}^{2^t-1} |j\rangle \otimes \left|a^{j_0}\right\rangle .$$

Next, after applying the controlled gate a^2 the state of the system will be

$$\phi_1 = \frac{1}{\sqrt{2^t}} \sum_{j=0}^{2^t-1} |j\rangle \otimes \left|a^{j_0+2j_1}\right\rangle .$$

Once all the operations in the controlled boxes are performed the final state of the system is

$$\begin{aligned}\phi_t &= \frac{1}{\sqrt{2^t}}\sum_{j=0}^{2^t-1} |j\rangle \otimes \left|a^{j_0+2j_1+\cdots+2^{t-1}j_{t-1}}\right\rangle \qquad (4.3)\\ &= \frac{1}{\sqrt{2^t}}\sum_{j=0}^{2^t-1} |j\rangle \otimes \left|a^{j}\right\rangle .\end{aligned}$$

Notice that this state is a balanced mixture containing all possible powers of a from 0 to $2^t - 1$.

Now, fix $b \in \mathbb{Z}_N^*$. In the sum (4.3) we can collect all the powers $a^j = b$, and we do this for each b. Henceforth, we may rewrite (4.3) as

$$\phi_t = \sum_{b\in\mathbb{Z}_N^*}\left(\frac{1}{\sqrt{2^t}}\sum_{j=0}^{2^t-1} x_{b,j}\,|j\rangle\right)\otimes|b\rangle ,$$

where $x_{b,j}$ is the binary sequence $\{x_{b,0}, \ldots x_{b,2^t-1}\}$ with $x_{b,j} = 1$ if $a^j = b$, and 0 otherwise.

At this stage we perform a measurement of the second register. Then the state collapses into

$$\psi_1 = \frac{1}{\sqrt{2^t}}\sum_{j=0}^{2^t-1} x_{b,j}\,|j\rangle$$

after detecting some b. It is not important what the element b is since all the valuable information is now encoded in the sequence $\{x_{b,j}\}$. The reason for this is that this sequence is clearly periodic with period r =order(a) because

$$a^j = a^{j+r} = \cdots a^{j+kr} = \cdots .$$

Thus, the QFT will allow us to discover this period r. After crossing the box QFT the sequence $\{x_{b,j}\}$ gets transformed into a new sequence of complex numbers $\{y_{b,0}, \ldots, y_{b,2^t-1}\}$ with larger norm $|y_{b,j}|^2$, if j is a multiple of the frequency $2^t/r$, and very small otherwise.

Hence, after we measure the first register we will get a zero or a one for each qubit. From this we read an integer $0 \leq k \leq 2^t - 1$. This k correspond to a value that is a multiple of $2^t/r$. That is,

$$k \approx l\frac{2^t}{r},$$

or equivalently

$$\frac{k}{2^t} \approx \frac{l}{r}.$$

The method of continued fraction expansions is used to compute a small denominator for the quotient $k/2^t$. The details can be read in [Chung], Chapter 5.

Chapter 5

Schrödinger Equation

5.1 Introduction

In this chapter we introduce infinite dimensional quantum systems, historically the first formalism of quantum mechanics.

Max Planck's quantization of energy for the black-body and Albert Einstein interpretation of Planck's quanta as photons, with energy proportional to their frequency, $E = hv$, where h is the Planck's constant, were amongst the first fundamental ideas.

Since energy and momentum are related in the same way as the frequency and the wave number, it was realized that the momentum of a photon was inversely proportional to its wavelength $p = h/\lambda$. De Broglie went further, and hypothesized that this was true for all particles. He showed that, assuming that the matter waves propagate along with their particle counterparts, electrons form standing waves, meaning that only certain discrete frequencies were allowed.

Elaborating on de Broglie's ideas, Schrödinger decided to find a proper 3-dimensional wave equation for the electron. The equation he proposed was the following:

$$i\hbar\frac{\partial\psi(r,t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla\psi(r,t) + V(r,t)\psi(r,t), \tag{5.1}$$

where (r,t) denote coordinates of position and time, and $\nabla\psi(r,t)$ is the Lapalacian. With respect to the spatial coordinates it is given by:

$$\nabla\psi(x,y,z,t) = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2},$$

where $V(r,t)$ is a potential function, m is the mass of the particle and $\hbar = h/(2\pi)$. We recall that the potential $V(r,t)$ is a real valued function such that the force exerted on the particle is given by

$$F(r,t) = -gradV(r,t).$$

The one dimensional wave equation would be

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial\psi^2}{\partial x^2} + V(x,t)\psi(x,t). \tag{5.2}$$

We will interpret this equation as the evolution equation of the state ψ. We will regard this function as an element of the Hilbert space $L^2(\mathbb{R}^3,\mathbb{C})$. That is, a complex valued function defined on $\mathbb{R}^3$ such that $\int\limits_{\mathbb{R}^3} |\psi(r,t)|^2\, dr < \infty$.

Since this is a finite integral we may assume that ψ is *normalized*, i.e., it has norm 1. The state ψ is interpreted as the probability to find the particle at time t within the cube $I \subset \mathbb{R}^3$ as

$$P = \int_I |\psi(r,t)|^2 \, dr.$$

We will take this equation as a axiom. However, we may give a heuristic argument to see why this equation is natural.

We know that the evolution equation of a state is given by the Hamiltonian

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi,$$

and that the Hamiltonian corresponds to the energy observable. In classical mechanics the Hamiltonian is given by

$$H = \frac{m}{2}v^2 + V = \frac{p^2}{2m} + V$$

the sum of the kinetic energy and the potential energy of the particle, where v denotes its velocity and $p = mv$ its momentum. As we shall discuss below, in one dimensions the correct analogue for the momentum will be the Hermitian operator given by

$$p = \frac{\hbar}{i}\frac{\partial}{\partial x}.$$

Hence, p^2 corresponds to the composition $p \circ p$

$$p^2 = \frac{\hbar}{i}\frac{\partial}{\partial x}(\frac{\hbar}{i}\frac{\partial}{\partial x}) = -\hbar^2 \frac{\partial^2}{\partial x^2},$$

and therefore the kinetic energy plus the potential energy will correspond to the operator

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V$$

and we see that (5.2) is just the evolution equation for the state, as we had deduced before.

To see why these are the correct analogue operators, remember that for an observable A and a state prepared as ψ the quantity

$$\langle A \rangle_\psi = \langle \psi, A\psi \rangle$$

gives the average of the measurement given by A. By using (5.2) we deduce that for a particle in state ψ its average position is given by

$$\langle x \rangle_\psi = \int_{\mathbb{R}} x \;\; |\psi(x,t)|^2 \, dx = \int_{\mathbb{R}} \overline{\psi} \; x \; \psi dx = \left\langle \overline{\psi}, x\psi \right\rangle.$$

Thus, *position* must be replaced by the operator *multiplication by* x: $\psi \to x\psi$.

What should be the operator corresponding to the particle's velocity? Well, the average velocity of the particle is given by

$$\begin{aligned} \frac{d\,\langle x \rangle_\psi}{dt} &= \frac{d}{dt}\int_{\mathbb{R}} \overline{\psi} \; x \; \psi dx = \int_{\mathbb{R}} \frac{\partial}{\partial t}(\overline{\psi} \; x \; \psi) dx \\ &= \int_{\mathbb{R}} x \left(\frac{\partial \overline{\psi}}{\partial t}\psi + \overline{\psi}\frac{\partial \psi}{\partial t} \right) dx. \end{aligned} \tag{5.3}$$

We use equation (5.2) (taking conjugates commutes with differentiation)

$$\begin{aligned}
\frac{\partial \psi}{\partial t} &= \frac{\hbar i}{2m}\frac{\partial \psi^2}{\partial x^2} - \frac{i}{\hbar}V\psi \\
\frac{\partial \overline{\psi}}{\partial t} &= -\frac{\hbar i}{2m}\frac{\partial \overline{\psi}^2}{\partial x^2} + \frac{i}{\hbar}V\overline{\psi}
\end{aligned}$$

to write (5.3) as

$$\begin{aligned}
\frac{d\left\langle x\right\rangle_\psi}{dt} &= \int\limits_{\mathbb{R}} x\left(\frac{\partial \overline{\psi}}{\partial t}\psi + \overline{\psi}\frac{\partial \psi}{\partial t}\right)dx \qquad (5.4)\\
&= \frac{\hbar i}{2m}\int\limits_{\mathbb{R}} x\left(\frac{\partial \overline{\psi}^2}{\partial x^2}\psi + \overline{\psi}\frac{\partial \psi^2}{\partial x^2}\right)dx \\
&= \frac{\hbar i}{2m}\int\limits_{\mathbb{R}} x\left(-\frac{\partial \overline{\psi}^2}{\partial x^2}\psi + \overline{\psi}\frac{\partial \psi^2}{\partial x^2}\right)dx \qquad (5.5)\\
&= \frac{\hbar i}{2m}\int\limits_{\mathbb{R}} x\frac{\partial}{\partial x}\left(\frac{\partial \psi}{\partial x}\overline{\psi} - \psi\frac{\partial \overline{\psi}}{\partial x}\right)dx.
\end{aligned}$$

We then use integration by parts to rewrite the last expression inside the integral. If we denote by W the expression $\frac{\partial \psi}{\partial x}\overline{\psi} - \psi\frac{\partial \overline{\psi}}{\partial x}$ we have

$$\int x\frac{\partial}{\partial x}W = xW - \int W$$

Thus,

$$\begin{aligned}
&\frac{\hbar i}{2m}\int\limits_{\mathbb{R}} x\frac{\partial}{\partial x}\left(\frac{\partial \psi}{\partial x}\overline{\psi} - \psi\frac{\partial \overline{\psi}}{\partial x}\right)dx \\
=\; &\frac{\hbar i}{2m}x\left.\left(\frac{\partial \psi}{\partial x}\overline{\psi} - \psi\frac{\partial \overline{\psi}}{\partial x}\right)\right|_{-\infty}^{\infty}dx - \frac{\hbar i}{2m}\int\limits_{\mathbb{R}}\left(\frac{\partial \psi}{\partial x}\overline{\psi} - \psi\frac{\partial \overline{\psi}}{\partial x}\right)dx \\
=\; &-\frac{\hbar i}{2m}\int\limits_{\mathbb{R}}\left(\frac{\partial \psi}{\partial x}\overline{\psi} - \psi\frac{\partial \overline{\psi}}{\partial x}\right)dx,
\end{aligned}$$

since $\lim_{x\to\infty}\left|\psi(x,t)\right| = 0$. Finally, we notice that

$$\int_{\mathbb{R}}\psi\frac{\partial \overline{\psi}}{\partial x} = \psi\overline{\psi}\Big|_{-\infty}^{\infty} - \int_{\mathbb{R}}\overline{\psi}\frac{\partial \psi}{\partial x}dx = -\int_{\mathbb{R}}\overline{\psi}\frac{\partial \psi}{\partial x}dx.$$

Therefore, from (5.5) we have:

$$\begin{aligned}
\frac{d\left\langle x\right\rangle_\psi}{dt} &= -\frac{\hbar i}{2m}\int\limits_{\mathbb{R}}\left(\frac{\partial \psi}{\partial x}\overline{\psi} - \psi\frac{\partial \overline{\psi}}{\partial x}\right)dx \\
&= -\frac{\hbar i}{2m}\left(\int_{\mathbb{R}}\overline{\psi}\frac{\partial \psi}{\partial x}dx + \int_{\mathbb{R}}\psi\frac{\partial \overline{\psi}}{\partial x}dx\right) \\
&= -\frac{\hbar i}{m}\int_{\mathbb{R}}\overline{\psi}\frac{\partial \psi}{\partial x}dx.
\end{aligned}$$

Thus, the expectation value of the momentum will be

$$\begin{aligned}\langle p\rangle_\psi &= m\frac{d\,\langle x\rangle_\psi}{dt} = -\hbar i\int_{\mathbb{R}}\overline{\psi}\frac{\partial\psi}{\partial x}dx\\ &= -\hbar i\left\langle \psi, \frac{\partial}{\partial x}\psi\right\rangle,\end{aligned}$$

and we may identify the momentum operator as derivation as

$$p = -\hbar i\frac{\partial}{\partial x},$$

as we had claimed above.

5.2 Time independent wave equation

For the rest of this chapter we will restrict ourselves to the one dimensional wave equation. In this section we deal with the case where the potential function only depends on x hence, independent of t

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial\psi^2}{\partial x^2} + V(x)\psi(x,t) \tag{5.6}$$

$$\phi(x,0) = f(x). \tag{5.7}$$

To solve this equation we use the method of separation of variables. That is, we look first for a solution of the form $\psi(x,t) = u(x)\phi(t)$. If we substitute ψ in (5.6) we obtain the equation

$$i\hbar\; u(x)\phi'(t) = -\frac{\hbar^2}{2m}u''(x)\phi(t) + V(x)u(x)\phi(t).$$

Dividing both sides by $u(x)\phi(t)$ one obtains

$$i\hbar\;\frac{\phi'(t)}{\phi(t)} = -\frac{\hbar^2}{2m}\frac{u''(x)}{u(x)} + V(x)$$

Since the left hand side depends on t only, and the right hand sides depends only on x, both sides must be equal to some constant E from which we obtain a pair of ordinary differential equations

$$\begin{aligned}\phi'(t) &= \frac{-i}{\hbar}E\phi(t)\\ \left(-\frac{\hbar^2}{2m}\frac{d}{dx^2} + V(x)\right)u(x) &= Eu(x).\end{aligned}$$

The first equation can be solved directly as

$$\phi(t) = Ae^{\frac{-i}{\hbar}Et},$$

for some constant A. For the second one we notice that the term inside the parenthesis is just the Hamiltonian operator H and therefore $u(x)$ corresponds to an eigenfunction of H, and E to its corresponding eigenvalue.

Suppose we know that these are discrete $E_1, E_2, \ldots, E_n, \ldots$, with corresponding eigenfunctions $u_1(x), u_2(x), \ldots, u_n(x), \ldots$. Thus, for each n we have a solution of (5.6) $\psi_n(x,t) = u_n(x)\phi(t)$. Each solution $\psi_n(x,t) = A_n u_n(x) e^{\frac{-i}{\hbar}E_n t}$ is called a *stationary solution.* The reason for this is clear: The probability distribution function does not change with time since

$$|\psi_n(x,t)|^2 = |A_n|^2 \, |u_n(x)|^2$$

does not depend on t.

There is no reason to expect that any of these solutions would satisfy the initial condition $\psi(x,0) = f(x)$. But since equation (5.6) is linear we may hope to achieve this by taking some linear combination of the solutions ψ_n. However, it can be rigorously shown that an "infinite linear combination" does work. That is, there is a solution of the form

$$\psi(x,t) = \sum_{i=0}^{\infty} c_n \psi_n(x,t) = \sum_{i=0}^{\infty} c_n u_n(x) e^{\frac{-i}{\hbar}E_n t},$$

where c_n are coefficients chosen so that

$$f(x) = \psi(x,0) = \sum_{i=0}^{\infty} c_n \psi_n(x,0).$$

Equivalently,

$$f(x) = \sum_{i=0}^{\infty} c_n u_n(x).$$

In many situations we can choose the eigenfunctions $u_1(x), u_2(x), \ldots, u_n(x)$ forming an *orthonormal* set. In this case one can always compute the coefficients c_n as

$$c_n = \langle u_n(x), f(x) \rangle = \int_{-\infty}^{\infty} \overline{u_n}(x) f(x) dx.$$

Let us discuss a specific example.

5.3 Infinite potential well

Suppose we have a particle of mass m confined to move in some interval $[0,a]$ such that

$$V(x) = \begin{cases} \infty, \text{ if } x < 0 \text{ or } x > a \\ 0 \qquad \text{if } 0 \leq x \leq a \end{cases}$$

We may imagine that V is the limit of potential energies that approach $+\infty$ when $x \to 0^+$ or when $x \to a^-$, and is very close to zero inside the interval $[0,a]$

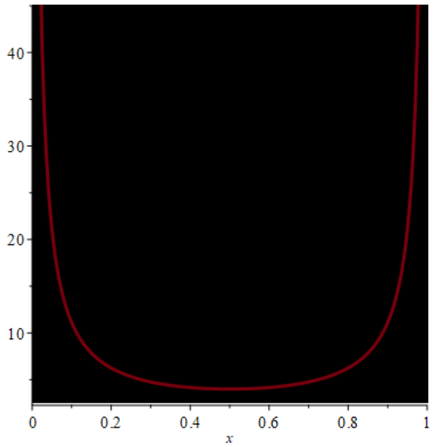


Figure 5.1: Potential Energy

In this case $u(x)$ must satisfy

$$u''(x) + \frac{2mE}{\hbar^2}u(x) = 0. \tag{5.8}$$

Since the potential is infinite outside the closed interval, the probability of finding the particle outside the *well* is zero, and therefore we may impose the boundary conditions

$$u(0) = u(a) = 0. \tag{5.9}$$

We first consider the cases $E = 0$ and $E < 0$. In both cases thee only solution to (5.8) satisfying (5.9) is trivial $u(x) = 0$. This, since $u(x) = Ax + B$ in the first case, where one can easily check that there are no trivial solutions. In the second case

$$u(x) = Ae^{-\sqrt{\frac{-2mE}{\hbar^2}}x} + Be^{\sqrt{\frac{-2mE}{\hbar^2}}x}$$

for suitable constants A and B. By substituting $x = 0$ one obtains $u(0) = 0 = A + B$; by substituting $x = a$ one gets

$$0 = Ae^{-\sqrt{\frac{-2mE}{\hbar^2}}a} + Be^{\sqrt{\frac{-2mE}{\hbar^2}}a}.$$

These two equations force $A = B = 0$.

Therefore, the only interesting case occurs when $E > 0$. Let $k = \sqrt{\frac{2mE}{\hbar^2}}$. In this case equation (5.8) has the general solution

$$u(x) = A\sin(kx) + B\cos(kx).$$

The condition $u(0) = 0$ implies that $B = 0$ and the condition $u(a) = 0$ implies that $\sin(ak) = 0$, which in turn implies that $k = n\pi/a$, $n = 1, 2, \ldots$. Hence, for each positive integer n one gets a solution

$$u_n(x) = \sin(\frac{n\pi x}{a}),$$

which is an eigenfunction of the Hamiltonian, with corresponding eigenvalue

$$E_n = \frac{k_n^2\hbar^2}{2m} = \frac{n^2\pi^2\hbar^2}{2ma^2}.$$

Now, each stationary solution $\psi_n(x,t) = A_n e^{\frac{-i}{\hbar}E_n t}\sin(\frac{n\pi x}{a})$ must be normalized hence we must choose A so that

$$\begin{aligned}\int_{-\infty}^{\infty} |A|^2 \left|e^{\frac{-i}{\hbar}Et}\right|^2 |u_n(x)|^2\, dx &= |A|^2 \int_0^a \sin^2(\frac{n\pi x}{a})dx = \\ &= |A|^2 \frac{a}{2} = 1.\end{aligned}$$

This forces $A = \sqrt{2/a}$.

One can readily check that the functions $u_n(x)$ are orthogonal since

$$\int_0^a \sin(\frac{n\pi x}{a})\sin(\frac{m\pi x}{a})dx = 0, \text{ if } m \neq n.$$

Thus, the general solution to the wave equation is given by

$$\psi(x,t) = \sqrt{\frac{2}{a}}\sum_{n=1}^{\infty} c_n \sin(\frac{n\pi x}{a})e^{\frac{-i}{\hbar}E_n t},$$

with

$$c_n = \int_{-\infty}^{\infty} \overline{u_n(x)}f(x)dx = \sqrt{\frac{2}{a}}\int_{-a}^{a} \sin(\frac{n\pi x}{a})f(x)dx.$$

As as particular example let us compute $\psi(x,t)$ when the initial condition is given by a function of the form $f(x) = Kx(x-a)$. The normalization condition forces $K = \sqrt{30/a^5}$. By integrating by parts one can compute c_n as:

$$c_n = \frac{-\sqrt{2}\sqrt{30}(n\pi\sin n\pi + 2\cos n\pi - 2)}{n^3\pi^3}.$$

But $\sin(n\pi) = 0$ and $\cos(n\pi) = (-1)^n$. Hence,

$$c_n = \begin{cases} \frac{8\sqrt{15}}{n^3\pi^3}, & \text{if } n \text{ is odd.} \\ 0 & \text{if } n \text{ is even} \end{cases}$$

Thus,

$$\psi(x,t) = \frac{8\sqrt{15}}{\pi^3}\sqrt{\frac{2}{a}}\sum_{n=1,3,5,7,\ldots}^{\infty} \frac{1}{n^3}\sin(\frac{n\pi x}{a})e^{\frac{-i}{\hbar}E_n t}, \text{ with } E_n = \frac{n^2\pi^2\hbar^2}{2ma^2}.$$

Remark 5.3.1.

1. Notice that even though each $\psi_n(x,t) = A_n e^{\frac{-i}{\hbar}E_n t}u_n(x)$ is a stationary solution, the linear combination of stationary solutions is, in general, not stationary. For instance, in the case where $\psi(x,t) = c_1\psi_n(x,t) + c_2\psi_m(x,t)$ and the coefficients c_i are real and the $u_n(x)$ are real functions (just for simplicity) one has

$$\begin{aligned}|c_1\psi_n(x,t) + c_2\psi_m(x,t)|^2 &= c_1^2|\psi_n(x,t)|^2 + c_2^2|\psi_m(x,t)|^2 + 2c_1c_2 Re\overline{\psi_n}(x,t)\psi_m(x,t) \\ &= c_1^2u_n^2 + c_2^2u_m^2 + 2c_1c_2u_n(x)u_m(x)\cos\left(\frac{E_m - E_n}{\hbar}t\right),\end{aligned}$$

an expression that varies periodically with t.

2. As a more specify example let us take an electron that moves in an infinite potential well in $[0,1]$. Its mass is equal to $m = 9.1 \times 10^{-31}$ kg. Recall $\hbar = 1.05 \times 10^{-34}$ J. s. We take $E_1 = \pi^2\hbar^2/2m = 5.9 \times 10^{-38}$ J. and $E_2 = 4\pi^2\hbar^2/2m = 2.4 \times 10^{-37}$ J. and

$$\psi(x,t) = \sqrt{2}\sin(\pi x)e^{\frac{-i}{\hbar}E_1 t} + \sqrt{2}\sin(4\pi x)e^{\frac{-i}{\hbar}E_2 t}.$$

This is a solution of the wave equation. The following two graphs correspond to $t = 0$ and $t = t_0$ such that $\cos\left(\frac{E_2 - E_1}{\hbar}t_0\right) = -1$

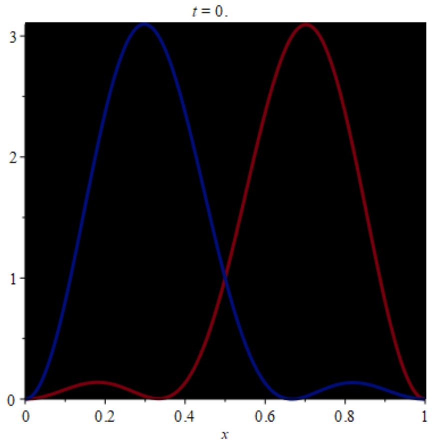


Figure 5.2: Evolution of a probability wave

3. We also notice that since the L^2 norm of $\psi(x,t) = \sum_{i=0}^{\infty} c_n u_n(x) e^{\frac{-i}{\hbar}E_n t}$ is equal to one, and the $u_n(x)$ are orthogonal, one has

$$1 = |\psi|_{L^2} = \sum_{i=0}^{\infty} |u_n|_{L^2} \left| e^{\frac{-i}{\hbar}E_n t} \right| |c_n|^2 = \sum_{i=0}^{\infty} |c_n|^2 .$$

Hence, with respect to the base $\{u_n(x)\}$ the measurement of the observable energy gives the value E_n with probability $|c_n|^2$.

5.4 Free particle

In this section we analyze the wave equation of a free particle that moves in one dimension. Hence, we assume that $V(x,t) = 0$ and therefore Schrödinger's equation becomes (as above)

$$\begin{aligned} i\hbar\frac{\partial\psi(x,t)}{\partial t} &= -\frac{\hbar^2}{2m}\frac{\partial\psi^2}{\partial x^2} \\ \psi(x,0) &= f(x). \end{aligned} \qquad (5.10)$$

Again, we apply the method of separation of variables and assume there is a solution of the form $\psi(x,t) = u(x)\phi(t)$. The difference with the case of a particle in a well is that we no longer have any boundary condition for $u(x)$. However, we still have as in (5.8) that

$$u''(x) + k^2 u(x) = 0, \text{ with } k = \sqrt{\frac{2mE}{\hbar^2}}.$$

The general (complex) solution of this equation is given by

$$u(x) = Ae^{ikx} + Be^{-ikx}.$$

On the other hand, as before,

$$\phi(t) = e^{-iEt/\hbar} = e^{-itk^2\hbar/2m}$$

and consequently

$$\psi(x,t) = Ae^{ik(x-tk\hbar/2m)} + Be^{-ik(x+tk\hbar/2m)}. \tag{5.11}$$

The function $\psi(x,t)$ is a linear combination of two waves. To understand the physical meaning of this equation, let us take first the first solution

$$\psi(x,t) = Ae^{ik(x-tk\hbar/2m)}.$$

As we recall from Chapter 1, a classical wave with period τ and wavelength λ that propagates at speed $v = \lambda/\tau$ in the positive direction of the x axis is described by a function of the form

$$q(x,t) = \cos(kx - \omega t), \text{ with } k = 2\pi/\lambda \text{ and } \omega = 2\pi/\tau.$$

The constant k is called the *wave number* and ω is called the *angular frequency.* The frequency (in Hertz) is $f = 1/\tau = \omega/(2\pi)$. We notice that v can also be written as $v = \omega/k$.

If we fix $t = t_0$, we may interpret $q(t_0, x)$ as a picture at time t_0 of an undulatory movement that propagates in the positive x direction at speed v. If $q(x,t)$ corresponds, for instance, to a circular wave of water in a pond, we would observe that the ripples "move" in the sense that if at a particular point at distance x the water forms a convex bump, next to its right we would observe a concave depression that will start rising as the bump decreases in height, and this effect propagates so that from peak to peak there is always a distance λ. If we fix x, on the other hand, water at this point will rise and fall cyclically with periodicity τ.

Now, if we look at the real part of $\psi(x,t)$

$$\psi(x,t) = Re\psi(x,t) = A\cos(kx - \frac{k^2\hbar}{2m}t)$$

this function represents a wave with angular frequency $\omega = k^2\hbar/(2m)$. Since $k = \sqrt{\frac{2mE}{\hbar^2}}$ we see that

$$E = \frac{k^2\hbar^2}{2m}. \tag{5.12}$$

According to wave-particle duality hypothesis, the de Broglie wavelength of a particle is inversely proportional to its momentum, or equivalently,

$$p = \hbar k = h/\lambda.$$

Equation (5.12) can be then written as $E = p^2/2m$, as expected.

We notice that none of the solutions (5.11) can be normalized. This corresponds to the fact that there no free particles with a definite energy. The solution to the wave equation in this case requires a weighted combination of waves of the form

$$\psi(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(k) e^{i(kx - \frac{\hbar k^2}{2m}t)} dk$$

where $g(k)$ is a suitable chosen function so that

$$f(x) = \psi(x,0) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} g(k) e^{ikx} dk.$$

This continuos form of the Fourier expansion is known as *Plancherel's theorem* (REF). The function $g(k)$ can be computed by taking the Fourier transform of $f(x)$

$$g(k) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} f(k) e^{-ikx} dk.$$

Let us analyze the concrete example of a free particle that is originally located inside the interval $[-a,a]$. That is, we assume that $\psi(x,0) = 1/\sqrt{2a}$ if $-a \le x \le a$, and $\psi(x,0) = 0$ outside this interval.

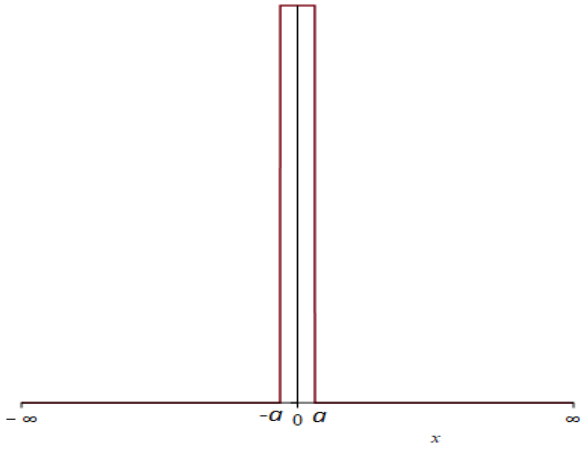


Figure 5.3

We calculate the function $g(k)$ as

$$\begin{aligned} g(k) &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2a}} \int\limits_{-a}^{a} e^{-ikx} dx = \frac{1}{\sqrt{2\pi a}} \left. \frac{e^{-ikx}}{-ik} \right|_{-a}^{a} \\ &= \frac{1}{k\sqrt{\pi a}} \frac{e^{ika} - e^{-ika}}{2i} = \frac{1}{\sqrt{\pi a}} \frac{\sin(ka)}{k}. \end{aligned}$$

Therefore,

$$\psi(x,t) = \frac{1}{\pi\sqrt{2a}} \int\limits_{-\infty}^{\infty} \frac{\sin(ka)}{k} e^{i(kx - \frac{\hbar k^2}{2m} t)} dk$$

This integral cannot be computed in terms of elementary functions. However, if a is small, we may approximate $\sin(ka) \approx ka$ and therefore we may write:

$$\psi(x,t) = \frac{1}{\pi}\sqrt{\frac{a}{2}} \int\limits_{-\infty}^{\infty} e^{i(kx - \frac{\hbar k^2}{2m} t)} dk.$$

Chapter 6

Random Computing

6.1 Introduction

Scientists continue to uncover ways in which nature manages and organizes itself. For example, the findings in the article [Hig21] have been described as "the first time we are seeing biology actively exploiting quantum effects."[1] Whereas mathematics is sometimes developed to study the physical world, there are times when the physical world provides a guide for machine development. In [Cu17], the authors cite the article [Ad94] as the foundational work on DNA computation and, furthermore, assert that DNA computing can be viewed as a means to obtain the first physical design of a non-deterministic universal Turing machine. We refer the interested reader to the discussion in section 3 of [Cu17] where the authors given an intriguing, albeit brief, description of molecular computing going back more than sixty years, beginning with [Fey60].

In this chapter we continue the theme of mathematically replicating the means by which nature can be seen as a computational engine. Specifically, we are interested in the manner in which certain aspects of diffusion are simultaneously, not sequentially, observable. For instance, we offer the following self-evident observation.

Assume that a heat source, such as a flame or a welding torch, is applied to the center of a circular disc of uniform thickness and material composition. Then two observers who are measuring temperatures at different points on the perimeter will detect a change of temperature at their points of contact at the same rate.

The diffusion in the above setting is that of heat. In what follows we study a diffusion process which admits a different visualization which we will call a *diffusion ring,* or simply a *ring.*

Within a circular ring, imagine a beam of light $\mathcal{B}$ (or some type of focused energy) emanating from a source at a perimeter point $\mathcal{P}_0$. Upon contact with another perimeter point $\mathcal{P}_1$ on the ring, the beam $\mathcal{B}$ splits into M sub-beams of equal magnitude in a prescribed set of directions toward perimeter points $\mathcal{P}_{2,1}, \cdots \mathcal{P}_{2,M}$. Let each sub-beam upon contact with some $\mathcal{P}_{2,k}$ split in manner as similar to the reflection of $\mathcal{B}$ at $\mathcal{P}_1$, and so on. Then after n such splittings, what portion of the original amount energy has returned to $\mathcal{P}_0$?

The analysis involves the mathematical understanding of the imagery of a beam reflecting and splitting within a diffusion ring. In this setting, we will count a single *diffusion step* as one instance

[1] *Bacteria Know How to Exploit Quantum Mechanics to Steer Energy,* SciTechDaily, March 29, 2021.

of contact, reflecting and subsequent splitting. To be precise, the count will be mathematically captured as one iteration of a symmetric matrix on a finite dimensional vector space. If each contact involves the splitting of a single beam into M sub-beams, then after n diffusion steps, one will have M^n paths of light traversing the ring. Though at this point we are solely interested in the mathematical aspects of our set-up, one cannot help but imagine the visualization of the diffusion ring. Indeed, if such a ring were 1 kilometer in diameter, and if the beam were to travel at the speed of light, then after 0.01 seconds one would expect to have more than M^{3000} sub-beams crossing various chords of the ring since in almost all circumstances more than 3000 diffusion steps would have taken place.

Having established our definition of a diffusion step, we now can state the first main result of this chapter.

Theorem 6.1.1. Let $N \geq 2$ be a positive integer which is assumed to be neither a prime nor a prime power. Let b denote an integer which is co-prime to N, and assume that the order of b modulo N is odd. Then the order of b modulo N can be computed in at most $O((\log N)^2)$ diffusion steps.

As it turns out, there is an effective bound for the number of diffusion steps in Theorem 6.1.1. Indeed, we prove that the number of diffusion steps n which are needed in Theorem 6.1.1 satisfies the bound

$$n < 4\log N(\lfloor \log_2 N \rfloor + 2), \tag{6.1}$$

where $\lfloor \cdot \rfloor$ is the floor function, log is the natural logarithm function, and $\log_2$ is the logarithm with base 2.

Our second result stems from an application of Theorem 6.1.1 to the problem of integer factorization. Again, we assume that N is neither a prime nor a prime power. So, there is an integer $m \geq 2$ such that

$$N = \prod_{i=1}^{m} p_i^{e_i}$$

where the primes $p_1, \cdots, p_m$ are distinct and the exponents $e_1, \cdots, e_k$ are strictly positive. Also, we will use the phrase *digital steps* to signify the term used to describe the computational complexity of an algorithm which is implemented on a classical Turing machine. With this, we can differentiate between *digital steps* and *diffusion steps* which quantify the complexity of algorithms which are carried out on either a digital computer or the above described diffusion machine.

Theorem 6.1.2. Let N be a positive integer with $m \geq 2$ distinct prime factors. Then with probability $p(m) \geq 1-(m+1)/2^m$ we can compute a non-trivial factor of N in at most $O((\log N)^2))$ digital steps and at most $O((\log N)^2)$ diffusion steps.

As with Theorem 6.1.1, the bound (6.1) applies for the number of diffusion steps in Theorem 6.1.2. It is entirely possible that the number of digital steps can be effectively bounded as well; however, for the sake of brevity, we choose not to study the effectiveness of the bound for the number of digital steps.

In case the algorithm terminates with no answer, one can simply repeat the computations. Under the usual assumptions of uniform random selection, if we execute the algorithm t times, then the probability of failing is less than

$$(1-p(m))^t = ((m+1)/2^m)^t \, .$$

In other words, the probability of success becomes arbitrarily close to 1 with sufficiently many implementations of the algorithm behind Theorem 6.1.2.

There exist deterministic algorithms which ascertain if N is either a prime or the power of a prime; see, for example, [AKS04], [Be07], or [Ra80]. In the two problems, the best known algorithms have (classical) complexity of order $O((\log N)^a)$ for some constant a. With this, we do not view the assumption that N is neither a prime nor a prime power as being restrictive, at least from the point of view of theoretical computability.

It is noteworthy that Shor's algorithm, which is the well-known method for factoring using a quantum computer, takes at most $O((\log N)^2 \log(\log N) \log(\log \log N))$ quantum steps. As with Shor's algorithm, Theorem 2 has a certain probability which is less than one of a successful completion. Initially, the probability of success for Shor's algorithm was determined to be at least $2/3$, and more recent studies have sought to optimize the probability of success; see, for example, [Za13]. Again, we will leave for elsewhere the problem of optimizing the probability of success of Theorem 6.1.2. In that regard, the methodology of [Za13] seems applicable.

There is current research into simulating Shor's algorithm on a digital computer; see, for example, [Mo16], [Pol09], [WHH17]. In that vein, we are able to readily simulate the diffusion computer behind Theorem 6.1.2, and we provide two examples. In the first, we take $N = 33$, and in the second we take $N = 1363$. The description of Theorem 6.1.2 for these examples is given below, and the computer code which was written in *Maple* is provided in an Appendix to this paper.

Our approach to proving Theorem 6.1.1 is as follows. Let r denote the order of b modulo N, which we write as $r = \text{ord}_N b$. Let V denote the set of powers of b modulo N, so the cardinality of V is r. As the notation suggests, the set V is viewed as the set of vertices of a graph. The edges of the graph are formed by connecting each b^k with points of the form b^{k2^j} where j ranges over positive and negative integers from $-(\lfloor \log_2 N \rfloor + 1)$ to $\lfloor \log_2 N \rfloor + 1$. We consider the diffusion process on the resulting graph associated to the so-called half-lazy random walk; see section 6.2.2 for details. From [Ba79] and [Lo75], we can express the eigenvalues of the associated Laplacian in terms of certain exponential sums. As it turns out, optimal bounds for these exponential bounds are known; see [KM12] as well as [Va09]. When combining these results, we show that after $O((\log N)^2)$ diffusion steps one determines the cardinality of V, thus the order of b modulo N.

As we will discuss, the bounds we employ for such exponential sums are worst-case scenarios. As such, we expect that in practice fewer diffusion steps may suffice to obtain some information about r.

Regarding Theorem 6.1.2, we follow the method which is used in Shor's algorithm and replace the quantum computation step with a diffusion computation. In doing so, it is necessary to choose an integer a whose order modulo N is even. For this, we prove how to reduce the problem, with sufficiently high probability, to an implementation of Theorem 6.1.1.

The outline of this paper is as follows. In section 2 we establish notation and recall necessary background material, including results from spectral graph theory and discrete time heat kernels, which are used elsewhere. Additionally, we formalize the concept of *diffusion process computing,* which stems from the ideas first presented in [HoRe20]. In section 3 we obtain results using modular arithmetic which are necessary in order to apply Theorem 6.1.1 to prove Theorem 6.1.2. The computations in section 3 appeal to known deterministic algorithms, such as the Euclidean algorithm. In section 4 we proof Theorem 1. Specifically, we construct the graph which is the mathematical realization of the above-described beam-splitting reflection ring. As stated, the bound on the number of diffusion steps in Theorem 6.1.1 is obtained by certain eigenvalue bounds which in this case are equivalent to bounds for exponential sums. In section 5 we combine the results from previous sections and complete the proof of Theorem 6.1.2. In section 6 we describe the digital implementation of Theorem 6.1.2 and obtain the factorization of $N = 33$ and $N = 1363$, and in section 7 we present a number of concluding remarks.

Finally, it is important to note that the ideas and methods of this paper were motivated by the Master's Degree thesis [HoRe20]. In [HoRe20] the phrase *heat computer* was coined, and the idea formed our motivation for a *diffusion computer.* Additionally, in [HoRe20] it is shown how to construct heat computers which solve Simon's problem and the Deutsch-Jozsa problem, and in each case the number of heat steps coincides with the number of quantum steps for the known quantum algorithm solutions of these problems. In other words, the conceptualization of a diffusion computer began with [HoRe20], and the work in this chapter can be considered as a furtherance of the initial ideas step forth in [HoRe20]

6.2 Preliminaries

6.2.1 Basic notation

Any graph X we consider in this chapter is finite, undirected and connected. The set of vertices V is finite, and the set of edges E consists of a collection of two-element subsets of V; we allow an edge to connect a vertex to itself, which may be called a self-loop. We let $k = |V|$ denote the number of vertices of X. Additionally, we assume X has a real-valued weight function $w\colon V \times V \to \mathbb{R}$ satisfying the following properties.

(i) **Symmetry:** For all $x, y \in V$, $w(x, y) = w(y, x)$.

(ii) **Semi-positivity:** For all $x, y \in V$, $w(x, y) \geq 0$.

(iii) **Positivity for Edges:** For all $x, y \in V$, $w(x, y) > 0$ if and only if $\{x, y\} \in E$.

The weight function generalizes the cases when there are multiple edges joining two vertices or self-loops.

Choose any ordering of the vertices. The corresponding adjacency matrix A of X is a $k \times k$ matrix whose (x, y)-entry is given by the weight function, meaning that $A(x, y) = w(x, y)$. From the above properties for the weight function, the adjacency matrix A is symmetric with non-negative real entries. The degree of a vertex $x \in V$ is defined as

$$d(x) = \sum_{y \in V} w(x, y).$$

A weighted graph X is said to be regular of degree d if $d(x) = d$ for all vertices $x \in V$.

6.2.2 Random walks and the discrete time heat kernel

We now assume that X is a regular weighted graph of degree d. A half-lazy random walk on X is a Markov chain with state space $(V, \mathcal{P}(V))$ with arbitrary initial probability distribution $p_0\colon V \to \mathbb{R}$, and transition probability matrix given by

$$W = \frac{1}{2}\left(I + \frac{1}{d}A\right). \tag{6.2}$$

The matrix W is called the half-lazy walk matrix of X. Intuitively, the half-lazy random walk is a process that starts with a single particle at some vertex, and at each step the particle either stays put at its current vertex with probability $1/2$ or moves randomly to a neighbor with probability $1/2$. In the second case, the particle moves from vertex x to vertex y with probability $w(x,y)/(2d)$.

Let $p_n\colon V \to \mathbb{R}$ denote the probability distribution at time n, meaning after n steps of the half-lazy random walk on X. Starting with an arbitrary probability distribution p_0 on V, then p_n is given inductively by

$$p_n(x) = \frac{1}{2}p_{n-1}(x) + \frac{1}{2}\sum_{y\in V}\frac{w(x,y)}{d}p_{n-1}(y).$$

Equivalently, we can view p_n as a column vector from $\mathbb{R}^k$, so this equation can be written in matrix form as

$$p_n = \frac{1}{2}\left(I + \frac{1}{d}A\right)p_{n-1} = Wp_{n-1} = W^n p_0.$$

Let us denote the (x,y)-entry of W^n by $w_n(x,y)$, which can be interpreted as the probability that a particle which follows the half-lazy random walk on X and starts at y is at vertex x after n steps.

The function $w_n(x,y)$ is called the discrete time heat kernel of X because the random walk provides a probabilistic interpretation of heat diffusion in X, where the temperature at a vertex x is considered to be a manifestation of "heat particles" which spread randomly in all directions. As such, one can view p_n as the distribution of these heat particles at time n. That is, if one starts with m_y units of heat at each vertex $y \in V$, then the temperature at vertex x after n steps is $p_n(x) = \sum_{y\in V} m_y w_n(x,y)$.

Under this interpretation, the physical principle of conservation of energy can be stated as

$$p_{n+1} - p_n = (W - I)p_n.$$

The difference $\partial_n p_n p_{n+1} - p_n$ is called the discrete time derivative, and the operator $\Delta W - I$ is referred to as the discrete Laplacian. With this notation the above equation reads

$$\partial_n p_n = \Delta p_n, \tag{6.3}$$

which is known as the discrete heat equation on X.

The standard solution of the discrete heat equation with initial condition p_0 is $p_n = W^n p_0$, and the solution can be expressed by diagonalizing the symmetric matrix W. Specifically, let $\lambda_0, \dots, \lambda_{k-1}$ be the eigenvalues of W with corresponding eigenvectors $\psi_0, \dots, \psi_{k-1}$ which form an orthonormal basis for $\mathbb{R}^k$. Then, using elementary linear algebra, one obtains that

$$p_n = \sum_{j=0}^{k-1}\langle\psi_j, p_n\rangle\psi_j = \sum_{j=0}^{k-1}\langle\psi_j, W^n p_0\rangle\psi_j = \sum_{j=0}^{k-1}\langle\psi_j, p_0\rangle\lambda_j^n\psi_j,$$

where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product of vectors in $\mathbb{R}^k$, meaning that if $\psi_j(x)$ denotes the x-entry of ψ_j, then

$$\langle \psi_j, p_0 \rangle = \sum_{x \in V} \psi_j(x) p_0(x).$$

In our notation, $\lambda_0 = 1$ and $|\lambda_j| < 1$ for all $j = 1, \cdots, k-1$. Also, $\psi_0(x) = 1/k$ for all $x \in V$. As a result, if n goes to infinity, the solution p_n converges to the uniform probability distribution on the set V regardless of the initial condition p_0. More precisely, we have the following proposition.

Proposition 6.2.1. With the notation as above, let λ_1 be the largest eigenvalue of W less than 1. Assume that the initial condition p_0 is a probability distribution, meaning it is semi-positive and has ℓ^1 norm equal to one. Then for all $x \in V$ and all $n \geq 0$ we have that

$$\left| p_n(x) - \frac{1}{k} \right| \leq \lambda_1^n.$$

Proof. From (6.2) we have that a vector ψ is an eigenvector of A with eigenvalue η if and only if ψ is an eigenvector of W with eigenvalue

$$\lambda = \frac{1}{2}\left(1 + \frac{1}{d}\eta\right).$$

The eigenvalues of the adjacency matrix A of a degree d regular weighted graph lie in the interval $[-d, d]$ with d being the largest eigenvalue; see for example Theorem 7.5 of [Ni18]. Hence the eigenvalues of W satisfy the inequalities

$$1 = \lambda_0 > \lambda_1 \geq \cdots \lambda_{k-1} \geq 0.$$

The eigenvector ψ_0 corresponding to λ_0 is such that $\psi_0(x) = 1/k$ for all $x \in V$. With this, we have for any $x \in V$ and all $n \geq 0$ the expansion

$$p_n(x) = W^n p_0(x) = \langle \psi_0, p_0 \rangle \psi_0(x) + \sum_{j=1}^{k-1} \langle \psi_j, p_0 \rangle \lambda_j^n \psi_j(x) = \frac{1}{k} + \sum_{j=1}^{k-1} \langle \psi_j, p_0 \rangle \lambda_j^n \psi_j(x), \tag{6.4}$$

where the last equality follows from the assumption that p_0 is a probability distribution.

Since the set of eigenvalues are orthonormal, we have a version of Parseval's formula, namely

$$\left| p_n(x) - \frac{1}{k} \right|^2 \leq \sum_{y \in V} \left| p_n(y) - \frac{1}{k} \right|^2 = \sum_{j=1}^{k-1} \left(\langle \psi_j, p_0 \rangle \lambda_j^n \right)^2 \leq \lambda_1^{2n} \sum_{j=1}^{k-1} \langle \psi_j, p_0 \rangle^2. \tag{6.5}$$

Moreover, by writing $p_0 = \sum\limits_{j=0}^{k-1} \langle \psi_j, p_0 \rangle \psi_j$ we obtain

$$\sum_{j=1}^{k-1} \langle \psi_j, p_0 \rangle^2 \leq \sum_{j=0}^{k-1} \langle \psi_j, p_0 \rangle^2 = \langle p_0, p_0 \rangle = \sum_{x \in V} p_0(x)^2 \leq \sum_{x \in V} p_0(x) = 1.$$

The last inequality follows from the fact that p_0 is a probability distribution on V. Combining this inequality with (6.5) we get

$$\left| p_n(x) - \frac{1}{k} \right|^2 \leq \lambda_1^{2n},$$

which completes the proof of the assertion. □

6.2.3 Weighted Cayley graphs of finite abelian groups

Let G be a finite abelian group, and let $S \subseteq G$ be a fixed symmetric subset generating G. The symmetry condition means that if $s \in S$ then $-s \in S$. Moreover, let $\alpha \colon S \to \mathbb{R}^{>0}$ be a function such that $\alpha(s) = \alpha(-s)$. One can construct a weighted Cayley graph $X = \mathrm{Cay}(G, S, \alpha)$ of G with respect to S and α as follows. The vertices of X are the elements of G. Two vertices x and y are connected with an edge if and only if $x - y \in S$. The weight $w(x, y)$ of the edge (x, y) is $w(x, y) = \alpha(x - y)$. With all this, one can show that X is a regular weighted graph of degree

$$d = \sum_{s \in S} \alpha(s).$$

By assuming that S generates G it follows that the Cayley graph X is connected.

Given $x \in G$, let χ_x denote the character of G corresponding to x; see, for example, [CR62]. Then χ_x can be represented as an eigenvector of the adjacency operator A of X with corresponding eigenvalue equal to

$$\eta_x = \sum_{s \in S} \alpha(s) \chi_x(s);$$

see Corollary 3.2 of [Ba79]. It follows that χ_x is an eigenvector of the half-lazy walk matrix W of X with eigenvalue

$$\lambda_x = \frac{1}{2}\left(1 + \frac{1}{d}\eta_x\right).$$

Let us number the *distinct* eigenvalues of W as

$$1 = \lambda_0 > \lambda_1 > \cdots > \lambda_l \geq 0,$$

where, obviously $1 \leq l \leq |G| - 1$.

The results of section 6.2.2 on half-lazy random walks on the graph X apply to deduce that for any initial probability distribution p_0 on the set G of vertices of X, the half-lazy random walk $\{p_n\}_{n=0}^{\infty}$ converges to the uniform distribution on G as $n \to \infty$. More precisely, for all $x \in G$ and all positive integers n, the inequality

$$\left| p_n(x) - \frac{1}{|G|} \right| \leq \lambda_1^n$$

holds true.

Moreover, fix $1 \leq i \leq l$, and let $E_i \subseteq \mathbb{R}^{|G|}$ be the eigenspace corresponding to the eigenvalue λ_i. The set $B_i = \{\chi_x : \lambda_x = \lambda_i\}$ is an orthonormal basis for E_i, so the projection h_i of p_0 onto the eigenspace E_i is given by

$$h_i = \sum_{\substack{x \in G \\ \lambda_x = \lambda_i}} \langle \chi_x, p_0 \rangle_G \, \chi_x,$$

where $\langle \cdot, \cdot \rangle_G$ denotes inner product on the set all complex-valued functions on G and is defined as

$$\langle f, g \rangle_G \colon = \frac{1}{|G|} \sum_{x \in G} \overline{f(x)} g(x).$$

With this in mind, the solution p_n of the discrete heat equation (6.3) on X subject to the initial probability distribution $p_0\colon G \to \mathbb{R}$ can also be written as

$$p_n = \sum_{i=0}^{l} \lambda_i^n \; h_i. \tag{6.6}$$

Expressing the solution in terms of the projections of the initial condition onto the eigenspaces of W has the advantage that it is not subject to a particular basis choice. As we shall see, this is particularly useful for defining the notion of a *Diffusion Computer.*

6.2.4 Diffusion process computing

A diffusion process in $X = \mathrm{Cay}(G, S, \alpha)$ may be regarded as an analog computation on X in the following sense.

Definition 6.2.2. A real-valued function h on G is said to be computable by a diffusion process in X with initial condition $p_0\colon G \to \mathbb{R}$ if the following holds. Let $\{p_m\}_{m=0}^{\infty}$ be the solution to the discrete heat equation (6.3) in X with initial probability distribution p_0. Then for any given $\varepsilon > 0$ there exists a positive integer $n = n(\varepsilon)$ such that for all $m > n(\epsilon)$ and all $x \in G$, we have that

$$|p_m(x) - h(x)| < \varepsilon.$$

Colloquially, we will refer to X as a *Heat Computer* or descriptively as a *Diffusion Computer.* The function h will be called *diffusion computable.* As stated, the concept of a Diffusion Computer first was developed in [HoRe20]. By the linearity of the discrete heat equation and the uniqueness of its solution, any linear combination of diffusion-computable functions on X is also computable by a diffusion process in X.

Theorem 6.2.3. Let f be an arbitrary real-valued function on G, and let $h_0, h_1 \ldots, h_l$ be the corresponding projections of f onto the eigenspaces $E_0, E_1 \ldots, E_l$ of the half-lazy walk matrix W. Then the functions $h_0, h_1 \ldots, h_l$ are diffusion computable on $X = \mathrm{Cay}(G, S, \alpha)$.

A formal proof is in [HoRe20]. The computation is carried out by induction and can be described in terms of the heat diffusion. First one computes h_0 by just letting heat flow for enough time and then reading the temperature of the steady-state solution. We repeat the same procedure, this time with initial temperature function $\lambda_1^{-n_*}(f - h_0)$ for a suitable value of n_*. The fact that the eigenvalues are strictly decreasing guarantees that the steady-state solution converges to h_1, and we can continue recursively in this manner.

It is important to note that in the Definition 6.2.2 above and throughout this chapter we use the expression “diffusion process" instead of the “heat computer" which was coined in [HoRe20]. We do so because the principle of computation stays the same if the heat is replaced by any other diffusion process.

Any physical device capable of implementing the recursive procedure described in Theorem 6.2.3 will be called a diffusion based computer, or simply a diffusion computer.

In essence, a diffusion computer is a device that computes the Fourier expansion of a complex-valued function f defined on G in its base of characters. The Fourier spectrum of f can be obtained as follows. Fix $1 \leq i \leq l$, and suppose that the eigenspace E_i has dimension m_i. Recall that

$$h_i(x) = \sum_{\substack{a \in G \\ \lambda_a = \lambda_i}} \langle \chi_a, f \rangle_G \, \chi_a(x), \quad x \in G.$$

Therefore, by first computing h_i and then evaluating it at m_i different elements of G we get a system of linear equations that allows us to solve for the Fourier coefficients $\langle \chi_a, f \rangle_G$ such that $\lambda_a = \lambda_i$. By letting i vary, we get the whole spectrum of f.

It is an extraordinary fact that each one of the iterations is carried out by an extremely simple repetitive procedure, namely that the value of the function at each vertex is replaced by the average value of its neighbors, and this same thing occurs at every vertex. In many cases the first projections are enough to infer interesting properties about f. For that matter, in this chapter we only exploit the capability of computing the zeroth projection of f.

6.2.5 Equating quantum steps and diffusion steps

In [Ma21], the author quotes the principal manager of the quantum-computing group at Microsoft Research in Redmond, Washington as saying that "Quantum computing is essentially matrix vector multiplication — it's linear algebra underneath the hood". From the linear algebra point of view, it is the opinion of the authors that, in general, a diffusion computer can be regarded as the ℓ^1 version of the quantum computer, which itself is based on ℓ^2 theory. Let us describe the meaning of this comment.

As it is known, a quantum computation entails two different types of operations. The first is a deterministic operation which is just the abstract version of the classical evolution equation in quantum mechanics. For this, a unitary vector ϕ in a Hilbert space $(\mathbb{C}^n, \ell^2)$ evolves into a new vector $\psi = U\phi$ where U is some unitary operator. The second operation involves a measurement of this new state. This procedure is non-deterministic and has the effect of "collapsing ψ". More precisely, each particular measurement is modeled by the decomposition of $\mathbb{C}^n$ into finite orthogonal subspaces H_i. By "collapsing" we mean composing ψ with the projection ψ_i onto H_i, and this projection occurs with probability $|\psi_i|^2$.

By comparison, on a diffusion computer the unitary vector ϕ is replaced by a stochastic vector (i.e. ℓ^1-norm one vector) in $(\mathbb{R}^N, \ell^1)$. The evolution of ϕ is determined by the symmetric operator W. Then it is reasonable to define one diffusion computation step as one application of W which maps ϕ to $W\phi$. A measurement, on the other hand, is just a classical inspection of the vertices of the graph X where the diffusion process takes place. It is evident that the inspection of a subset of elements of X is also a projection operation. As with a quantum computer, a projection also is counted as one step.

In summary, one determines the steps in a quantum computation by counting the number of compositions of unitary matrices and projections, and in a diffusion process by counting the number of compositions of symmetric matrices and projections. As such, we consider a count of quantum steps to be comparable to a count of diffusion steps.

6.3 Some number theoretic considerations

As before, we let N be a positive, odd integer which we write as a product

$$N = \prod_{i=1}^{m} p_i^{e_i} \tag{6.7}$$

with $m \geq 2$ different odd prime factors with exponents $e_i > 0$. As such, we assume that N is not prime and not a prime power. Let $\mathbb{Z}_N$ denote the set of inequivalent classes of integers modulo N. There is a natural mapping

$$\mathbb{Z}_N \to \mathbb{Z}_{p_1^{e_1}} \times \cdots \times \mathbb{Z}_{p_m^{e_m}} \tag{6.8}$$

defined by

$$a \mapsto (a \bmod p_1^{e_1}, \cdots, a \bmod p_m^{e_m}).$$

By the classical Chinese Remainder Theorem, (6.8) is an isomorphism. Furthermore, the mapping (6.8) yields an isomorphism of the respective commutative rings, which we write as

$$g_N \colon \mathbb{Z}_N^* \to \mathbb{Z}_{p_1^{e_1}}^* \times \cdots \times \mathbb{Z}_{p_m^{e_m}}^*. \tag{6.9}$$

In a slight abuse of notation, we occasionally use x to denote either an element in $\mathbb{Z}_N^*$ or its image $g_N(x)$. For each i, the ring $\mathbb{Z}_{p_i^{e_i}}^*$ is a cyclic group under multiplication. Let us denote its generator by u_i. We shall write the order of u_i as

$$\operatorname{ord}_{p_i^{e_i}} u_i = p_i^{e_i-1}(p_i - 1) = 2^{c_i} p_i'$$

where $c_i > 0$ and p_i' is an odd positive integer.

Without loss of generality we assume that the primes $p_1, \ldots, p_m$ are ordered so that we have the inequalities $c_1 \geq c_2 \geq \cdots \geq c_m$.

The following proposition computes, under certain circumstances, a non-trivial square root of 1 modulo N.

Proposition 6.3.1. For any $a \in \mathbb{Z}_N^*$, let $s \geq 0$ be the least power of 2 such that

$$a^{2^s q} \equiv 1 \bmod N$$

for some odd integer q. Let us write

$$g_N(a) = (u_1^{d_1}, \ldots, u_m^{d_m}).$$

If there exist i, j with $1 \leq i < j \leq m$ such that d_i is odd and d_j is even, then $s > 0$. Furthermore, if we set $x = a^{2^{s-1}q}$ then

$$x^2 \equiv 1 \bmod N \quad \textit{and} \quad x \not\equiv \pm 1 \bmod N.$$

Proof. By assuming d_i is odd and d_j is even, we can write $d_j = 2^v d_j'$ where $v > 0$ and d_j' is an odd integer.

Because $u_i^{2^s q d_i} \equiv 1$ in $\mathbb{Z}_{p_i^{e_i}}^*$ the order of u_i, which is $2^{c_i} p_i'$, divides $2^s q d_i$. This implies that $s \geq c_i > 0$ and p_i' divides $q d_i$. Thus $s \geq 1$. Set

$$x = a^{2^{s-1}q} = (u_1^{2^{s-1}q d_1}, \ldots, u_m^{2^{s-1}q d_m}).$$

Then

$$x^2 = (u_1^{2^s q d_1}, \ldots, u_m^{2^s q d_m}) \equiv 1 \bmod N$$

Moreover, by the minimality of s, we have that $x \not\equiv 1 \bmod N$. It remains to show that $x \not\equiv -1 \bmod N$.

If $x \equiv -1 \bmod N$, then for every k we have that $x \equiv -1 \bmod p_k^{e_k}$. Therefore, if $x \not\equiv -1 \bmod p_k^{e_k}$ for a single index k, then $x \not\equiv -1 \bmod N$. Indeed, we now will prove that $x \not\equiv -1 \bmod p_j^{e_j}$ for the particular index j for which d_j is odd and $j > i$, which is assumed to exist as stated in the premise of the proposition. More specifically, we claim that $x \equiv 1 \bmod p_j^{e_j}$, which we rewrite as

$$u_j^{2^{s-1} q d_j} \equiv u_j^{2^{s+v-1} q d'_j} \equiv 1 \bmod p_j^{e_j}. \tag{6.10}$$

Equation (6.10) holds if and only if $2^{c_j} p'_j$, the order of u_j divides $2^{s+v-1} q d'_j$. Equivalently, equation (6.10) holds if and only if

$$s + v - 1 \geq c_j \quad \text{and} \quad p'_j \mid q d'_j. \tag{6.11}$$

Because $u_j^{2^s q d_j} \equiv 1 \bmod p_j^{e_j}$, the order of u_j, which is $2^{c_j} p'_j$, divides $2^s q d_j = 2^{s+v} q d'_j$. Since all integers p'_j, q and d'_j are odd, we then have that $s + v \geq c_j$ and $p'_j \mid q d'_j$. Thus, the second condition in (6.11) is proved. Since $s \geq c_i \geq c_j$ and $v > 0$, we have that

$$s + v - 1 \geq c_i + v - 1 \geq c_j,$$

which proves the first condition in (6.11) and completes the proof of the proposition. □

Lemma 6.3.2. Let N be a product of $m \geq 2$ distinct odd primes, and let $M = \lfloor \log_2 N \rfloor + 1$. For any $a \in \mathbb{Z}_N^*$, define the subset $S = S(a)$ of $\mathbb{Z}_N^*$ by

$$S = \{a^{\pm 2^t} \ \bmod \ N : \text{for all } \ t = 0, \ldots, M\}.$$

If there are repetitions in S, then with probability $p(m) = 1 - (m+1)/2^m$ we can find an $x \in \mathbb{Z}_N^*$ in at most $O(\log_2 N)$ deterministic steps for which

$$x^2 \equiv 1 \bmod N \quad \textit{and} \quad x \not\equiv \pm 1 \bmod N. \tag{6.12}$$

Proof. Let us assume there is at least one repetition in S, meaning that for some ℓ and ℓ' we have that $a^{2^l} \equiv a^{\pm 2^{l'}} \bmod N$. Without loss of generality, we may assume that $l > l'$. In other words, we have that

$$a^{2^l \pm 2^{l'}} \equiv 1 \bmod N. \tag{6.13}$$

Solving this equation, we obtain

$$a^{2^{l'}(2^{l-l'} \pm 1)} \equiv 1 \bmod N \quad \text{so then} \quad a^{2^{l'} q} \equiv 1 \bmod N,$$

where $q = 2^{l-l'} \pm 1$, which is an odd integer.

Once the set S has been constructed, and once one is given the values of ℓ and ℓ' in (6.13), then one can determine the smallest s such that $a^{2^s q} \equiv 1 \bmod N$ in at most $O(\log_2 N)$ deterministic computations.

As stated, the mapping (6.9) is a bijection between $\mathbb{Z}_N^*$ and $\mathbb{Z}_{p_1^{e_1}}^* \times \cdots \times \mathbb{Z}_{p_m^{e_m}}^*$. Let $\{u_1, \ldots, u_m\}$ denote a set of generators of the cyclic multiplicative groups $\mathbb{Z}_{p_1^{e_1}}^*, \ldots, \mathbb{Z}_{p_m^{e_m}}^*$. With this, we conclude that any element $a \in \mathbb{Z}_N^*$ is uniquely determined by the set of exponents $(d_1, \ldots, d_m)$ for which $g_N(a) \equiv (u_1^{d_1}, \ldots, u_m^{d_m})$. In other words, choosing a random element $a \in \mathbb{Z}_N^*$ is equivalent to choosing a random m-tuple $(d_1, \ldots, d_m)$, where each d_i is randomly and uniformly chosen from the set $\{1, \ldots, p_i^{e_i-1}(p_i - 1)\}$, $i = 1, \ldots, m$. Note that the set of integers from 1 to $p_i^{e_i-1}(p_i - 1)$ has equal number of even and odd elements.

With the above discussion, we can now estimate the probability $p(m)$. Proposition 6.3.1 proves that when $s > 0$ and if we set $x = a^{2^{s-1}q}$, then (6.12) holds when d_i is odd and d_j is even for some $1 \leq i < j \leq m$. Assuming a is chosen randomly and uniformly, let us show that the probability that a satisfies the conditions on i and j, hence d_i and d_j, by showing that the probability of the complementary event is $(m + 1)/2^m$.

The complementary event consists of the following $m + 1$ mutually exclusive events: First, there is an integer k with $1 \leq i \leq m$ such that for $i \leq k$ each d_i is even and for all $j > k$ each d_j is odd; and, second, all values of d_k are odd. Each of these events has probability $1/2^m$, hence the union has probability $(m + 1)/2^m$, so then $p(m)$ has probability $1 - (m + 1)/2^m$, as claimed. □

Remark 6.3.3. In the statement of Lemma 6.3.2, we assumed that we were given the two elements in S which form a repetition, meaning they are equivalent $\mod N$. In other words, the analysis in Lemma 6.3.2 begins with (6.13). Since S has $O(\log_2 N)$ elements, the straightforward, exhaustive search in S to determine if (6.13) occurs takes at most $O((\log_2 N)^2)$ deterministic steps.

Lemma 6.3.4. *Let G be a finite cyclic group of even order n, which we write as $n = 2^c m$ where $c > 0$ and m is an odd integer. Let u be a generator of G. Then for $1 \leq d \leq 2^c m$ we have the following statements.*

1. *If d is odd, then $\mathrm{ord}_G(u^d) = 2^c k'$, where k' is odd.*
2. *If d is even, which we write as $d = 2^v d'$ with d' odd, then $\mathrm{ord}_G(u^d) = 2^{c-t} m'$, where $t = \min(c, v) > 0$ and m' is odd.*

Proof. The assertion follows immediately from the elementary observation that $\mathrm{ord}_G(u^d) = n/\gcd(n, d)$. □

Lemma 6.3.5. *For any $a \in \mathbb{Z}_N^*$, the largest power of 2 that divides $\mathrm{ord}_N(a)$ is less than $\log_2 N$. Therefore, if $M = [\log_2 N] + 1$, then the order of $b = a^{2^M}$ must be odd.*

Proof. As above, let $g_N(a) = (u_1^{d_1}, \ldots, u_m^{d_m})$, from which we have that

$$\mathrm{ord}_N(a) = \mathrm{lcm}(\mathrm{ord}_{p_1}(u_1^{d_1}), \ldots, \mathrm{ord}_{p_m}(u_m^{d_m})) = \mathrm{lcm}(2^{c_1} p_1', \ldots, 2^{c_m} p_m').$$

Then by Lemma 6.3.4 the largest power of 2 that divides $\mathrm{ord}_N(a)$ is less than or equal to $\max\{c_1, \ldots, c_m\}$ which is necessarily less than $\log_2 N$ because

$$\log_2 N > \log_2 \left(\prod_{i=1}^{m} 2^{c_i} p_i' \right) = \sum_{i=1}^{m} (c_i + \log_2 p_i') > \max\{c_1, \ldots, c_m\}.$$

Let $r = \mathrm{ord}_N(a)$, and write $r = 2^v r'$ where r' is odd. Since $r < N$, it follows that $v < M$. Let $\mathrm{ord}_N(a^{2^M}) = 2^e t$ where t is odd and $e \geq 0$. Then $r \mid 2^{M+e}t$; consequently, $r' \mid t$, so we can factor t as $t = r't'$ where t' is odd. Then

$$a^{2^M t} = a^{2^v 2^{M-v} r't'} = (a^{2^v r'})^{2^{M-v} t'} \equiv 1 \bmod N.$$

Thus, $\mathrm{ord}_N(a^{2^M})$ divides t, which is odd, so then $\mathrm{ord}_N(a^{2^M})$ itself is odd, as claimed. □

Proposition 6.3.6. Set $M = [\log_2 N] + 1$. For any $a \in \mathbb{Z}_N^*$, let $r_b = \mathrm{ord}_N(a^{2^M})$. If r_b is known, we can compute $r_a = \mathrm{ord}_N(a)$ in at most $O(\log_2 N)$ deterministic steps. Furthermore, r_a is even with probability at least $p(m) = 1 - (m+1)/2^m$ in which case $x = a^{r_a/2}$ satisfies

$$x^2 \equiv 1 \bmod N \quad \textit{and} \quad x \not\equiv \pm 1 \bmod N.$$

Proof. By Lemma 6.3.5, r_b is odd. Then $\mathrm{ord}_N(a)$ is equal to $2^k r_b$, where k is the smallest exponent such that $a^{2^k r_b} \equiv 1 \bmod N$. Necessarily, we have that $k \leq M$. Given r_b, we can compute such k in at most $O(\log_2 N)$ steps. Let $g_N(a) = (u_1^{d_1}, \ldots, u_m^{d_m})$. Proposition 6.3.1 shows that $k > 0$, i.e. $\mathrm{ord}_N(a)$ is even if there exist i, j with $1 \leq i < j \leq m$ such that d_i is odd and d_j is even . According to the proof of Lemma 6.3.2, the probability of this is at least $p(m)$. The statement now follows from Proposition 6.3.1. □

6.4 Proof of Theorem 6.1.1

Let $N \geq 2$ be a fixed positive integer, and let b be an integer which is relatively prime to N and of *odd* order r in $\mathbb{Z}_N$. The proof that we can determine r in at most $O((\log_2 N)^2)$ diffusion steps runs as follows.

1. We construct an appropriate weighted Cayley graph $X_{r,S}$ with r vertices and study the half-lazy random walk $\{p_k^{X_{r,S}}\}_{k=0}^{\infty}$ on this graph in n steps. The construction of the graph is undertaken without the explicit knowledge of r.

2. By using the bounds for the Korobov-type exponential sums as proved in [KM12], we will deduce an upper bound for the second largest eigenvalue $\lambda_*^{X_{r,S}}$ of the half-lazy random walk on $X_{r,S}$. We note here that results related to those from [KM12] are given in [LLW98], [Mo09] and [Va09].

3. We determine the number of diffusion steps n so that $\lambda_*^{X_{r,S}} < N^{-2}$. In doing so, the integer r can be characterized as the integer closest to $1/p_n^{X_{r,S}}(e)$, where $e = 0$ is the starting vertex of the graph $X_{r,S}$ in additive notation.

As it will be shown, n can be taken to be the smallest integer bigger than $4 \log N(\lfloor \log_2 N \rfloor + 2)$.

6.4.1 Graph construction in multiplicative notation

For a fixed $b \in \mathbb{Z}_N^*$ of odd order r, let $G_{N,b} = \langle b \rangle \subseteq \mathbb{Z}_N^*$ be the subgroup of $\mathbb{Z}_N^*$ generated by b. The elements of $G_{N,b}$ are the classes of $b^k \bmod N$ for $k = 0, 1, \ldots, r-1$. Take $S_{N,b} = \{b^{\pm 2^t} : t = 0, \ldots, M\}$ with $M = \lfloor \log_2 N \rfloor + 1$. Define the weight function $\alpha_{N,b}$ by

$$\alpha_{N,b}(b^{2^t}) := |\{l \in \{0, \ldots, M\} \colon b^{2^t} \equiv b^{2^l} \bmod N \ \text{ or } \ b^{2^t} \equiv b^{-2^l} \bmod N\}|,$$

where $|A|$ denotes the number of elements of a finite set A. It is immediate that $\alpha_{N,b}(b^{2^t}) = \alpha_{N,b}(b^{-2^t})$, for $t = 0, \ldots, M$. Therefore, $\alpha_{N,b}$ can be used as the weight function in the construction of the weighted Cayley graph in the multiplicative notation, as in [Ba79].

Let $X_{N,b} = \mathrm{Cay}(G_{N,b}, S_{N,b}, \alpha_{N,b})$ be the weighted Cayley graph of $G_{N,b}$ with respect to $S_{N,b}$ and the weight function $\alpha_{N,b}$. It is immediate that $X_{N,b}$ has r vertices and it is a regular graph of degree $2(M+1)$.

It is important to note that in the construction of $X_{N,b}$ we do not know the value of r. Indeed, all that is required is the value of N since we begin with one point b and, recursively, let the diffusion process develop in $2(M+1)$ possible directions from any given point.

From the beginning, we do not know the entire graph. Nevertheless, since diffusion is local in nature, this allows us to build $X_{N,b}$ one diffusion step at a time. Our main theorem states that after a number of steps which is polynomial in $\log_2 N$ we will have enough information to approximate the number of vertices of $X_{N,b}$, and thus the order of b. What makes the process effective is the fact that diffusion occurs simultaneously at all constructed vertices, which provides some form of parallel computation.

6.4.2 An equivalent description of the graph

In this section we will describe the graph $X_{N,b}$ in additive notation, which we find to be more amenable for describing the bounds for the eigenvalues. Since b is fixed, we will simplify the notation by omitting the index b and emphasizing the dependence upon r.

It is important to emphasize that, ultimately, we will use diffusion to compute r. The only information we will use about r itself is that $0 < r \leq N$ and that r is the order of an element b modulo N. Nonetheless, this section is provided as a notational aid in our proofs of Theorem 6.1.1 and Theorem 6.1.2.

Let $C_r = \{0, \ldots, r-1\}$ denote the additive group of integers modulo r, and set

$$S = \{\pm 2^j : j = 0, \ldots, M\} \quad \text{with} \quad M = \lfloor \log_2 N \rfloor + 1.$$

Note that $S \subseteq C_r$ is a symmetric set which generates C_r, and that the numbers $\pm 2^j$ for $j \in \{0, \ldots, M\}$ are not necessarily distinct modulo r. We define the weight function $\alpha \colon S \to \mathbb{R}^{>0}$ as follows. For $2^j \in S$, we let

$$\alpha(2^j) := |\{l \in \{0, \ldots, M\} \colon 2^j \equiv 2^l \bmod r \ \text{ or } \ 2^j \equiv -2^l \bmod r\}|.$$

The congruence $\pm 2^j \equiv -2^l \bmod r$ is equivalent to $\pm 2^j \equiv 2^l \bmod r$. Hence, $\alpha(-2^j) = \alpha(2^j)$, so it can be used as the weight function in the construction of the weighted Cayley graph with respect to S.

Since b has order r modulo N, the weight function α is equal to the weight function $\alpha_{N,b}$. As a result, the values of α can be determined by computing powers of b modulo N where the value of r is unknown.

Let $X_{r,S} = \mathrm{Cay}(C_r, S, \alpha)$ be the weighted Cayley graph of C_r with respect to S and α. The graph $X_{r,S}$ has r vertices, and it is regular of degree $|S| = 2(M+1)$. In case all numbers $\pm 2^j$ with $j \in \{0, \dots, M\}$ are distinct modulo r, the graph $X_{r,S}$ is the Cayley graph of C_r with respect to S. If there are repetitions modulo r in the sequence $\pm 2^j$ with $j \in \{0, \dots, M\}$, then we can view $X_{r,S}$ in such a way that the vertex $x \in C_r$ is connected to the vertex $y \in C_r$ with possibly more than one edge; the number of edges connecting x and y being the number of elements of the set $\{l \in \{0, \dots, M\} \colon x - y \equiv 2^l \mod r \text{ or } x - y \equiv -2^l \mod r\}$.

For this choice of S, the eigenvalues η_k for $k = 0, \dots, r-1$ of the adjacency matrix for the graph are known. Specifically, $\eta_0 = 2(M+1)$ and

$$\eta_k = \sum_{x \in S} \alpha(x) e^{\frac{2\pi i}{r} kx} = \sum_{j=0}^{M} e^{\frac{2\pi i}{r} k 2^j} + \sum_{j=0}^{M} e^{-\frac{2\pi i}{r} k 2^j} \quad \text{for each} \quad 1 \le k \le r-1.$$

Thus, the eigenvalues of the half-lazy walk matrix $W_{r,S}$ of $X_{r,S}$ are

$$\lambda_k^{X_{r,S}} = \frac{1}{2}\left(1 + \frac{\eta_k}{2(M+1)}\right) \quad \text{for} \quad k = 0, 1, \dots, r-1.$$

Proposition 6.2.1 implies that for any vertex x we have

$$\left| p_n^{X_{r,S}}(x) - \frac{1}{r} \right| \le (\lambda_*^{X_{r,S}})^n, \tag{6.14}$$

where, as above, $\lambda_*^{X_{r,S}}$ is the largest eigenvalue of $W_{r,S}$ less than 1.

6.4.3 Eigenvalue bounds

Our next task is to find an upper bound independent of r for the quantity

$$\frac{1}{2(M+1)} \max_{k \in \{1, \dots r-1\}} |\eta_k| = \frac{1}{2(M+1)} \max_{k \in \{1, \dots r-1\}} \left| \sum_{j=0}^{M} e^{\frac{2\pi i}{r} k 2^j} + \sum_{j=0}^{M} e^{-\frac{2\pi i}{r} k 2^j} \right|. \tag{6.15}$$

Let $k \in \{1, \dots, r-1\}$ be arbitrary, and let $d = \gcd(k, r)$. Set $k_1 = k/d$ and $r_1 = r/d$. Then, we have that

$$\sum_{j=0}^{M} e^{\frac{2\pi i}{r} k \cdot 2^j} = \sum_{j=0}^{M} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j}. \tag{6.16}$$

Since r_1 and k_1 are relatively prime, $e^{\frac{2\pi i}{r_1} k_1}$ is a primitive r_1-th root of unity. Let $\tau_1 = \mathrm{ord}_{r_1} 2$. In order to bound (6.16), we consider the following three cases.

(i) **Assume** $r_1 \ge 4$ **and** $M \ge \tau_1 - 1$**.** Since $M + 1 \ge \tau_1$, we can write $M + 1 = q\tau_1 + s$ for $q \ge 1$ and some $0 \le s \le \tau_1 - 1$. Equivalently, $M = q\tau_1 + s_1$, for $q \ge 1$ and $-1 \le s_1 \le \tau_1 - 2$. Then, we have that

$$\sum_{j=0}^{M} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j} = q \sum_{j=0}^{\tau_1 - 1} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j} + \sum_{j=0}^{s_1} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j},$$

where, in the case that $s_1 = -1$ the sum on the right-hand side is taken to be zero. The second statement of Corollary 1 of [KM12], in our notation, states that when $r_1 > 3$ we have the bound

$$\max_{(k_1,r_1)=1} \left| \sum_{j=1}^{\tau_1} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j} \right| < \tau_1 - 1,$$

where the maximum is taken over all pairs of coprime k_1 and r_1. Since $\tau_1 = \mathrm{ord}_{r_1} 2$, it is immediate that

$$\sum_{j=1}^{\tau_1} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j} = \sum_{j=0}^{\tau_1 - 1} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j}.$$

Therefore,

$$\begin{aligned} \left| \sum_{j=0}^{M} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j} \right| &\leq q \left| \sum_{j=0}^{\tau_1 - 1} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j} \right| + \left| \sum_{j=0}^{s_1} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j} \right| \\ &\leq q(\tau_1 - 1) + s_1 + 1 \\ &= M + 1 - q \leq M. \end{aligned}$$

(ii) **Assume** $r_1 \geq 4$ **and** $M < \tau_1 - 1$**.** Since r_1 is odd, we actually have that $r_1 \geq 5$. In this case we proceed analogously as in the proof of [KM12], Theorem 2. This approach is justified, because in the notation of [KM12], $\mathcal{L} = \lfloor \log_2 r_1 \rfloor \leq \lfloor \log_2 r \rfloor \leq \lfloor \log_2 N \rfloor = M - 1$, so there are sufficiently many summands in the series so that the results of [KM12], Lemma 3, apply.

More precisely, Lemma 3 from [KM12], claims that if $(a, q) = 1$, $q > 3$, then for every integer $m \geq 0$ there exist an integer ℓ with $m < \ell \leq \lfloor \log_2 q \rfloor + 1 + m$ and such that

$$\left| \exp\left(\frac{2\pi i}{q} a 2^{\ell} \right) - 1 \right| < \left| \exp\left(\frac{2\pi i}{q} a 2^{\ell - 1} \right) - 1 \right|.$$

Taking $q = r_1 \geq 5$, $a = k_1$ and $m = 0$, we conclude that there exist an integer $\ell_0 \in \{1, \ldots, M\}$ such that

$$\left| \exp\left(\frac{2\pi i}{r_1} k_1 2^{\ell_0} \right) - 1 \right| < \left| \exp\left(\frac{2\pi i}{r_1} k_1 2^{\ell_0 - 1} \right) - 1 \right|.$$

Therefore,

$$\begin{aligned} \left| \exp\left(\frac{2\pi i}{r_1} k_1 2^{\ell_0 - 1} \right) + \exp\left(\frac{2\pi i}{r_1} k_1 2^{\ell_0} \right) \right| &\leq \left| \exp\left(\frac{2\pi i}{r_1} k_1 2^{\ell_0 - 1} \right) + 1 \right| \\ &= \frac{\left| \exp\left(\frac{2\pi i}{r_1} k_1 2^{\ell_0} \right) - 1 \right|}{\left| \exp\left(\frac{2\pi i}{r_1} k_1 2^{\ell_0 - 1} \right) - 1 \right|} < 1. \end{aligned}$$

This shows that

$$\begin{aligned} \left| \sum_{j=0}^{M} e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j} \right| &= \left| \exp\left(\frac{2\pi i}{r_1} k_1 2^{\ell_0 - 1} \right) + \exp\left(\frac{2\pi i}{r_1} k_1 2^{\ell_0} \right) \right| \\ &\quad + \sum_{j \in \{0, \ldots, M\} \setminus \{\ell_0 - 1, \ell_0\}} \left| e^{\frac{2\pi i}{r_1} k_1 \cdot 2^j} \right| < 1 + M - 1 = M. \end{aligned} \tag{6.17}$$

(iii) **Assume** $r_1 = 3$**.** In this case $r = 3d$. Since k_1 and r_1 are relatively prime, $k_1 \in \{1, 2\}$, so then $k = k_1 d < 3d$. Therefore, we need to find the upper bound for the absolute value of the sums

$$\sum_{j=0}^{M} e^{\frac{2\pi i}{3} \cdot 2^j} \quad \text{or} \quad \sum_{j=0}^{M} e^{\frac{4\pi i}{3} \cdot 2^j}.$$

Trivially, both sums are bounded by M. Indeed, all terms in either sum are cube roots of unity. In both case, the terms corresponding to $j = 0$ and $j = 1$ add to give $1/2 \pm i\sqrt{3}/2$, respectively, which has absolute value equal to one. The remaining $M - 1$ terms have absolutely value equal to one, from which the bound of M for each series is obtained.

Since $e^{-\frac{2\pi i}{r} k \cdot 2^j} = e^{\frac{2\pi i}{r}(r-k) \cdot 2^j}$, the inequality (6.17) holds true for the negative exponents as well. Hence the bound for (6.15) becomes

$$\frac{1}{2(M+1)} \max_{k \in \{1, \ldots r-1\}} |\eta_k| < \frac{2M}{2(M+1)} = 1 - \frac{1}{M+1}.$$

In summary, we have proved that the largest eigenvalue of the half-lazy random walk matrix $W_{r,S}$ on $G_{r,S}$ satisfies the bound

$$\lambda_*^{X_{r,S}} \leq \frac{1}{2}\left(1 + 1 - \frac{1}{M+1}\right) = 1 - \frac{1}{2(M+1)}. \tag{6.18}$$

6.4.4 Counting the number of diffusion steps

It is now left to establish the number of diffusion steps needed to determine r by performing the half-lazy random walk on the weighted graph $X_{r,S}$, constructed above, by starting at vertex $e = 0$.

By combining (6.18) with (6.14) we deduce, for any positive integer n, that

$$\left| p_n^{X_{r,S}}(e) - \frac{1}{r} \right| \leq \left(1 - \frac{1}{2(M+1)}\right)^n.$$

We know that $p_n^{X_{r,S}}(e)$ converges to $1/r$ as n tends to infinity. For any two distinct positive integers m_1 and m_2 less than N, the smallest distance between their reciprocals $1/m_1$ and $1/m_2$ is bounded from below by

$$\frac{1}{N} - \frac{1}{N-1} = \frac{1}{N(N-1)} \geq \frac{1}{N^2}.$$

Thus, if we have $p_n^{X_{r,S}}(e) = 1/r$ within an error of $1/N^2$, we have determined r. Therefore, the smallest number of diffusion steps needed to determine r is bounded from above by the smallest positive integer n for which

$$\left(1 - \frac{1}{2(M+1)}\right)^n < \frac{1}{N^2}.$$

Since $N \geq 2$, we have that $0 < 1/(2M) < 1$ so then

$$-\log\left(1 - \frac{1}{2(M+1)}\right) = \sum_{\ell=1}^{\infty} \frac{1}{\ell(2(M+1))^\ell} > \frac{1}{2(M+1)}.$$

Choose n to be the smallest integer for which

$$n > 4(M+1)\log N = 2\log N(\lfloor \log_2 N \rfloor + 2).$$

Then, the inequality

$$\left| p_n^{X_{r,S}}(e) - \frac{1}{r} \right| < \frac{1}{N^2}$$

holds true, and then we can compute the integer r uniquely after n diffusion steps, followed by one measurement (of the “heat" at the starting point e).

6.5 Proof of Theorem 6.1.2

Our algorithm receives as input a positive integer N which is assumed to be neither prime nor the power of a prime. The algorithm returns a divisor d of N, with probability at least $p(m)$, and it runs as follows:

```
1. Select $a \in \mathbb{Z}_N = \{1, \ldots, N\}$ uniformly at random.
   Compute $d = \gcd(a, N)$.
       If $1 < d < N$, return $d$.
       Else go to step 2.

2. Compute the set $S = \{a^{\pm 2^t} \bmod N : t = 0, \ldots, M\}$ where $M = \lfloor \log_2 N \rfloor + 1$.

3. If there are repetitions in S, say $a^{2^l} = a^{\pm 2^{l'}}$ with $l > l'$, do:
       Set $q = 2^{l-l'} \pm 1$.
       Compute the smallest integer $s \geq 0$ such that $a^{2^s q} \equiv 1 \bmod N$.
       If $s > 0$, compute $d = \gcd(a^{2^{s-1}q} - 1, N)$
           If $1 < d < N$, return $d$.
           Else terminate (with no answer).
       Else terminate (with no answer).
   Else go to step 4.

4. Set $b = a^{2^M}$.
   Run the diffusion computer algorithm to determine the order of $b$ modulo $N$.
   Set $r_b = \mathrm{ord}_N(b)$.

5. Compute the smallest integer $k \geq 0$ such that $a^{2^k r_b} \equiv 1 \bmod N$.
   Set $r_a = 2^k r_b$ which is the order of $a$ modulo $N$.
   If $r_a$ is even, compute $d = \gcd(a^{r_a/2} - 1, N)$
       If $1 < d < N$, return $d$.
       Else terminate (with no answer).
   Else terminate (with no answer).
```

The Euclidean algorithm determines the greatest common divisor of two numbers in $\mathbb{Z}_N$ using at most $O((\log N)^2)$ deterministic steps. With this, we can bound the complexity of each of the above steps as follows.

Step 1: This steps uses the Euclidean algorithm to find $\gcd(a, N)$, so it runs in deterministic time $O((\log N)^2)$.

Step 2: Using the method of repeated squaring, we can compute the set S in at most $O(2M) = O(\lfloor \log_2 N \rfloor + 1)$ deterministic steps.

Step 3: Checking for repetitions in S requires $O(\log_2 N)$ steps. By Lemma 6.3.2, we can compute the smallest integer $s \geq 0$ such that $a^{2^s q} \equiv 1 \bmod N$ in $O(\log_2 N)$ deterministic steps. If $s > 0$, computing $\gcd(a^{2^{s-1}q} - 1, N)$ takes $O((\log N)^2)$ more deterministic steps.

Step 4: Theorem 6.1.1 states that using a diffusion computer allows us to compute r_b in at most $O((\log_2 N)^2)$ diffusion steps.

Step 5: By Lemma 6.3.6, we can compute r_a, the order of a, in $O(\log_2 N)$ deterministic steps. If r_a is even, computing $\gcd(a^{r_a/2} - 1, N)$ takes $O((\log N)^2)$ more deterministic steps.

The algorithm runs Steps 1 and 2, and then runs either Step 3 or Steps 4 and 5. Therefore, the algorithm runs in at most $O((\log N)^2) + O(\lfloor \log_2 N \rfloor + 1) + O(\log_2 N) + O((\log N)^2)$ determistic steps plus at most $O((\log_2 N)^2)$ diffusion steps.

The algorithm is successful if it returns a nontrivial factor d of N. Both Steps 3 and 5 have a success probability of at least $p(m) = 1 - (m+1)/2^m$, by Lemmas 6.3.2 and 6.3.6. Hence, whether the algorithm runs Step 3 or Steps 4 and 5, the success probability is at least $p(m)$.

6.6 Examples

The following examples illustrate the above described algorithm.

6.6.1 Example 1: $N = 33$

Step 1. We choose $a = 5$ which is relatively prime to 33.

Step 2. $M = \lfloor \log_2(N) \rfloor + 1 = 6$, and $S = \{5^{2^0} = 5, 5^{2^1} \equiv 25, \dots, 5^{2^5} \equiv 25, \dots\} \bmod 33$.

Step 3. $5^{2^5 - 2^1} \equiv 1 \bmod 33$, so then $s = 1$ and $d = \gcd(5^{2^0 \cdot 15} - 1, 33) = 11$.

Therefore, 11 divides N, from which we obtain that $N = 3 \times 11$.

6.6.2 Example 2: $N = 1363$

Step 1. We choose $a = 991$ which is relatively prime to 1363.

Step 2. $M = \lfloor \log_2(N) \rfloor + 1 = 11$, and $S = \{991^{2^0} = 991, 991^{2^1} = 721, \dots, 991^{2^{11}} = 944, \dots\} \bmod 1363$. There are no repetitions in S. Therefore we go to Step 4.

Step 4. We set $b = 991^{2^{11}} \equiv 944 \bmod 1363$ and check for repetitions in the set $S = \{b^{\pm 2^t} : t = 0, \dots, 11\}$ finding none. Thus, we run the diffusion computer in order to determine the order r_b of $b = 944$.

Note: In order to get an estimate for $1/r_b$ with an error less than $1/N^2 \approx 5.38 \times 10^{-7}$, the diffusion computer requires at least $\lfloor 4(M+1)\log(N) \rfloor + 1 = 347$ diffusion steps. As we will describe below, after 36 diffusion steps, consisting of $n = 25$ iterations of the diffusion process and 11 measurements, we were able t conclude that $r_b = 161$.

Step 5. The smallest non negative integer k such that $991^{2^k\times 161} \equiv 1 \bmod 1363$ turns out to be 1. We conclude that $r_a = 322$. So then we computed $\gcd(991^{161} - 1, 1363) = 47$. Thus, 47 divides N, from which we obtain that $N = 47 \times 29$.

Note that since 47 and 29 are prime and odd, then r_a must divide $(47 - 1)(29 - 1) = 1288$. Indeed, we have that, as expected, $1288 = 322 \cdot 4$.

Let us now discuss the details behind Step 4. After iterating the diffusion process $n = 25$ times, we then measured the probability distribution $p_{25}(v)$ for the 11 values of v which are the vertices on the graph corresponding to S. As it turns out, for each such v we had that $1/160 < p_{25}(v) < 1/162$. This narrows the possibilities for the order of b to just three integer values. By trying every one of these values we then confirmed that, indeed, $r_b = 161$.

Though not needed, the *Maple* code in the Appendix produces the values of $p_{25}(v)$ for all v. A histogram for the reciprocals of these probabilities is presented in the next Figure

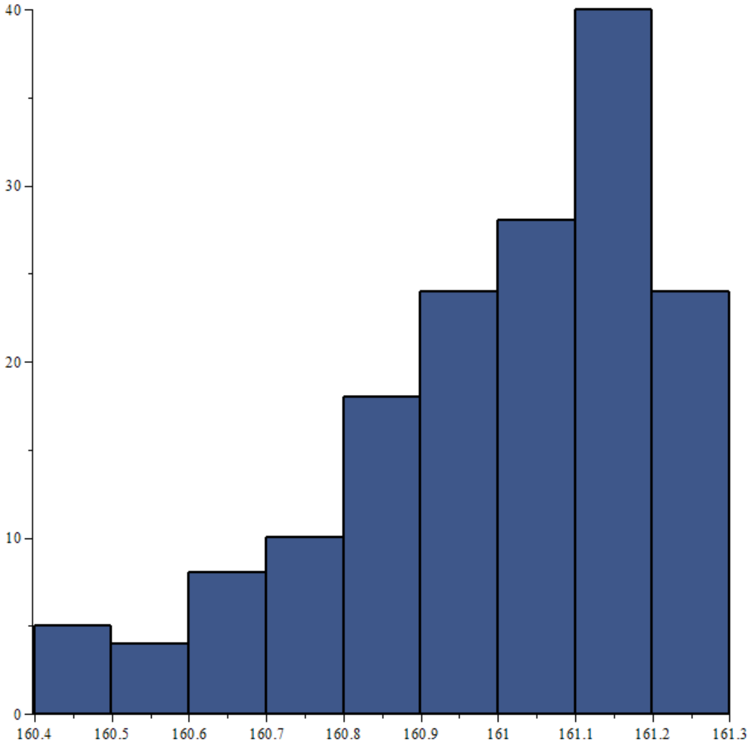


Figure 6.1: Interference pattern

In the next section we discuss a way in which the methodology of Example 2 can be encoded, possibly yielding a faster algorithm.

6.7 Concluding remarks

6.7.1 Reducing the number of diffusion steps

Note that the bound in Proposition 6.2.1 is somewhat crude since it used inequality (6.5). More specifically, it is possible that there are additional cancellations in the exponential sums appearing in the spectral expansion of the heat kernel for the half-lazy random walk. If so, then one would need fewer than the number of diffusion steps given by (6.1) in Theorem 6.1.1. The discussion provides a guide by which the implementation of the diffusion computation may yield results before reaching the number of steps stated in (6.1).

Let us suppose that after m diffusion steps we have that

$$\left|p_m(v) - \frac{1}{r}\right| \leq A_m \tag{6.19}$$

for any vertex v and some constant A_m. This is equivalent to

$$p_m(v) - A_m \leq \frac{1}{r} \leq p_m(v) + A_m.$$

Trivially, one has that $1/N \leq 1/r \leq 1$. Let $\mathcal{S}$ denote any subset of the vertices of X. We now can conclude that

$$\max\left(1/N, \max_{v\in\mathcal{S}} \left(p_m(v) - A_m\right)\right) \leq \frac{1}{r} \leq \min\left(1, \min_{v\in\mathcal{S}} \left(p_m(v) + A_m\right)\right). \tag{6.20}$$

Proposition 6.2.1 shows that the worst-case scenario happens when $A_m = \lambda_1^m$. However, it may be the case that additional cancellations occur in the spectral expansion (6.4) and in the exponential sums defining λ_1.

In general, let us suppose that, for whatever reason, after m diffusion steps we have the bound (6.19) for some A_m. Then the order r is amongst the integers h such that $1/h$ lies in the interval

$$\mathcal{I}_{\mathcal{S}} := \left(\max\left(1/N, \max_{v\in\mathcal{S}} \left(p_m(v) - A_m\right)\right), \min\left(1, \min_{v\in\mathcal{S}} \left(p_m(v) + A_m\right)\right)\right) \bigcap \{1/h \in \mathbf{Q} : h \in \mathbf{Z}\}. \tag{6.21}$$

Theorem 6.1.1 amounts to saying that by taking $m = \lfloor 4(M+1)\log(N)\rfloor + 1 = O((\log N)^2)$, there is only one integer, namely the order r, such that $1/r$ lies in the interval (6.21).

Certainly, for smaller m, and in the presence of an improved bound for A_m, the interval (6.21) could contain a small number of entries. One could stop the implementation of further diffusion steps and then check one by one each entry of $\mathcal{I}_{\mathcal{S}}$ to determine which value yields the sought-for order. In other words, one can reduce the number of iterations of the symmetric matrix W from the bound stated in (6.1) to m, but then increase the number of measurements from one to $|\mathcal{I}_{\mathcal{S}}|$, where $|\mathcal{I}_{\mathcal{S}}|$ is the cardinality of $\mathcal{I}_{\mathcal{S}}$. If it takes t steps on a digital computer to determine if a given integer h is the sought-for order, one would increase the number of digital steps by $t \cdot |\mathcal{I}_{\mathcal{S}}|$. For example, if for some constant $c > 0$ we take

$$m = 2\log((\log N)/c)(\lfloor \log_2 N\rfloor + 2),$$

then, as in section 6.4.4, we get that

$$A_m \leq \left(1 - \frac{1}{2(M+1)}\right)^m \leq \left(\frac{c}{(\log N)}\right)^2.$$

Let us rewrite the interval in (6.20) as

$$L_m \leq 1/r \leq U_m.$$

then

$$|\mathcal{I}_{\mathcal{S}}| \leq \frac{1}{L_m} - \frac{1}{U_m} = \frac{U_m - L_m}{L_m U_m}$$

and $|\mathcal{I}_{\mathcal{S}}|$ could be as small as $O((\log N)^k)$ for some constant $k \geq 2$.

In other words, there may be circumstances under which the number of diffusion steps could be reduced to $O(\log N)$, with an effective constant, while the number of digital steps would become $O((\log N)^k)$, thus not significantly changing the complexity of the number of digital steps. This discussion leads to an interesting optimization problem, we which will leave for a future study.

6.7.2 Searching for additional repetitions

The algorithm which proves Theorem 6.1.2 has, as Step 3, a check for repetitions amongst the set S only at the initial construction of the graph. However, this point could be exploited further at future diffusion steps, and the detection of such a repetition would detect the order r. For example, if the order of the element is 561, then by writing $561 = 512 + 32 + 16 + 1$, we will have a repetition on the fourth diffusion step. Also, if the order of the element is 111, then since $111 = 128 - 16 - 1$, one would have a repetition on the second diffusion step. In general, there will be a repetition after $n + 1$ steps if and only if one has $a^k \equiv a^\ell \mod N$ for distinct integers k and ℓ each of whose binary expansion consists of h or fewer non-zero digits.

In Step 3 of Theorem 6.1.2, we used $O(\log N)$ digital steps to detect repetitions. This investigation occurred at the first diffusion step. The naive method to test for further repetitions at diffusion step h would require $O((\log N)^h)$ digital steps. At this time, we have not devised a diffusion algorithm which would more efficiently undertake the problem of seeking higher repetitions.

6.7.3 Diffusion solutions of the Simon and Deutsch-Jozsa problems

For each integer $h \geq 2$, the Hamming cube, or hypercube graph, Q_h of dimension h is constructed as follows. The vertices correspond to h-tuples where each entry is either 0 or 1. There are 2^h vertices. An edge is formed by connecting two vertices v_1 and v_2 if and only if v_1 and v_2 differ in one and only one place. There are $2^{h-1}h$ edges.

In [HoRe20] the discrete time heat kernel $p_n^{(h)}(v)$ on Q_h is studied, and its spectral expansion is explicitly stated. With this information, Simon's problem and the Deutsch-Jozsa problem are studied in Chapter 4 of [HoRe20]. To recall, the statement of these problems are as follows.

1. **The Deutsch-Jozsa problem.** Let f be a Boolean function on the vertices of Q_h, meaning the range of values of f is $\{0, 1\}$. Suppose we know that f is either constant or balanced, by which we mean that the pre-image of either value 0 or 1 is one-half of the vertices of Q_h. Determine the number of steps required to decide if f is constant or balanced.

2. **Simon's problem.** Let f be a real-valued function on the vertices of Q_h. Let us view v_1 and v_2 as vectors of integers mod 2. Assume there is a vertex s, meaning an h-tuple mod 2, such that $f(v_1) = f(v_2)$ if and only $v_1 + v_2 \equiv 0 \bmod 2$ or $v_1 + v_2 \equiv 0 \mod 2$. Determine the number of steps required to compute s.

Using a digital computer, the Deutsch-Jozsa problem requires, in the worst case, $2^{h-1} + 1$ evaluations to be solved. On a quantum computer, the solution is obtained after a single quantum step. It is shown in section 4.4.1 of [HoRe20] that a *heat computer* constructed from discrete time heat diffusion on Q_h provides a heat computer solution to the Deutsch-Jozsa problem in a single heat step, which means the same as a diffusion step.

Using a probabilistic approach, Simon's problem can be solved on a digital computer in $O(2^{h/2})$ steps, whereas on a quantum computer a solution is obtained in $O(h)$ quantum steps. Similarly, it is shown in section 4.4.2 of [HoRe20] that a heat computer can solve Simon's problem in $O(h)$ diffusion steps.

We find it fascinating that a diffusion computer, or heat computer as the concept was called in [HoRe20], can efficiently solve the Deutsch-Jozsa problem and Simon's problem. In addition, the number of diffusion steps in the solutions coincide with the number of quantum steps needed to answer the questions using a quantum computer. In that regard, Theorem 6.1.2 can be compared to Shor's algorithm.

Bibliography

[AD] Aspect, A., Dalibard, J., & Roger, G. (1982). Experimental test of Bell's inequalities using time-varying analyzers. Physical review letters, 49(25), 1804.

[Ad94] Adelman, L.: *Molecular computation of solutions to combinatorial problems*, Science **266** (1994), 1021–1024.

[AKS04] Agrawal, M., Kayal, N., Saxena, N.: *Primes in P*, Annals of Math **160** (2004), 781–793.

[Ah01] Aharonov, D., Ambainis, A., Kempe, J., Vazirani, U.: *Quantum walks on graphs*, in: STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing (2001) Pages 50—59.

[Ba79] Babai, L.: *Spectra of Cayley graphs*, J. Combin. Theory Ser. B **27** (1979), no. 2, 180–189.

[Bell] Bell's Theorem: The naive view of an experimentalist, http://arxiv.org/ftp/quant-ph/papers/0402/0402001.pdf

[Benn 2014] Bennett, C. H., & Brassard, G. (2014). Quantum cryptography: Public key distribution and coin tossing. Theor. Comput. Sci., 560(P1), 7-11.

[Be07] Bernstein, D.J., Lenstra, H.W., Pila, J.: *Detecting perfect powers by factoring into coprimes*, J. Math. Comput. **76(257)** (2007), 385–388.

[BV97] Bernstein, E., Vazirani, U.: *Quantum complexity theory*, SIAM Journal on Computing **26(5)** (1997) 1411–1473.

[Cu17] Currin, A., Korovin, K., Ababi, M., Roper, K., Kell, D., Day, P., King, R.: *Computing exponentially faster: implementing a non-deterministic universal Turing machine using DNA*, Journal of The Royal Society Interface **14(120)** (2017), 20160990.

[C-T] Cohen-Tannoudji: Quantum Mechanics Vol I (2002).

[CR62] Curtis, C.W., Reiner, I.: *Representation theory of finite groups and associative algebras*, Reprint of the 1962 original. Wiley Classics Library. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1988.

[Chung] Chung, I. L., and M. A. Nielsen. "Quantum Computing and Quantum Information." (2000).

[Eraser] https://en.wikipedia.org/wiki/Quantum_eraser_experiment

[Fey 2011] Feynman, R. P., Leighton, R. B., & Sands, M. (2011). The Feynman lectures on physics, Vol. III.

[Fey 2010] Feynman, R. P., Hibbs, A. R., & Styer, D. F. (2010). Quantum mechanics and path integrals. Courier Corporation.

[Fey60] Feynman, R.: *There's plenty of room at the bottom*, Caltech. Eng. Sci. **23** (1960), 22–36.

[Gr 2005] Griffiths, D. J. (2005). Introduction to quantum mechanics. Pearson Education India.

[Gr] Griffiths D. Introduction to Quantum Mechanics.

[Hig21] Higgins, J., Lawson, T., Lloyd, S., Sohail, S., Allodi, M., Otto, J., Saer, R., Wood, R., Massey, S., Ting, P-C., Blankenship, R., Enge, G.: *Photosynthesis tunes quantum-mechanical mixing of electronic and vibrational states to steer exciton energy transfer*, Proc. Natl. Acad. Sci., **118(11)** (2021), e2018240118.

[HoRe20] Hoyos Restrepo, P.: *On the discrete heat equation and Kolmogorov complexity theory*, MSc thesis, Universidad Nacional de Colombia, Medellín, Colombia, 2020.

[KM12] Kaczorowski, J., Molteni, G. : *Extremal values for the sum* $\sum_{r=1}^{\tau} e(a2^r/q)$, J. Number Theory **132** (2012), 2595–2603.

[KoM10] Kondo, S., Miura, T.: *Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation*, Science **329** (2010), 1616–1620.

[LLW98] Liu, J., Liu, M., Wang, T.: *The number of powers of 2 in a representation of large even integers II*, Sci. China Ser. A-Math. **41** (1998), article no. 1255.

[LB] Le Bellac, M. (2011). Quantum physics. Cambridge University Press.

[Lo75] Lovász, L.: *Spectra of graphs with transitive groups*, Period. Math. Hungar. **6** (1975), no. 2, 191–195.

[vMW12] van Melkebeek, D., Watson, T.: *Time-space efficient simulations of quantum computations*, Theory Comput. **8** (2012), 1–51.

[Mar18] Marletto, C., Coles, D., Farrow, T., Vedral, V.: *Entanglement between living bacteria and quantized light witnessed by Rabi splitting*, Journal of Physics Communications, **2** (2018), No. 101001

[Ma21] Matthews, D.: *How to get started in quantum computing*, Nature **591** 7848 (2021).

[Mo09] Molteni, G.: *Cancellation in a short exponential sum*, J. Number Theory **130** (2010), 2011–2027.

[Mo16] Monz, T., Nigg, D., MArtinez, E., Brandl, M., Schinder, P., Rines, R., Wang, S., Chuang, I., Blatt, R.: *Realization of a scalable Shor algorithm*, Science **351**, (2016) 1068–1070.

[Ni18] Nica, B.: *A brief introduction to spectral graph theory*, EMS Textbooks in Mathematics, European Mathematical Society, Zurich, 2018.

[Pol09] Politi, A., Matthews, J., O'Brien, J.: *Shor's quantum factoring algorithm on a photonic chip*, Science **325** (2009), no. 5945, 1221.

[Ra80] Rabin, M.O.: *Probabilistic algorithm for testing primality*, J. Number Theory **12(1)** (1980), 128–138.

[RSA] R. Rivest, A. Shamir, L. Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems.

[Than13] Thanou, D., Dong, X., Kressner, D., Frossard, P.: *Learning heat diffusion graphs*, IEEE Trans. Signal Inform. Process. Netw. 3 (2017), no. 3, 484–499.

[TCM] Cadavid C, Hoyos, P. , Jorgenson J., Smajlović L, Vélez J.,An integer factorization algorithm which uses diffusion as a computational engine (Preprint).

[TA] Terras, Audrey. Fourier analysis on finite groups and applications. No. 43. Cambridge University Press, 1999.

[Va09] Vandehey, J.: *Differencing methods for Korobov-type exponential sums*, J. Anal. Math. **138** (2019), 405–439.

[WHH17] Wang, D. Hill, C., Hollenberger, L.: *Simulations of Shor's algorithm using matrix product states*, Quantum Inf. Process. **16** (2017), no. 7, Paper No. 176, 13 pp.

[Wa12] Watrous, J.: *Quantum computational complexity*, in: Computational complexity. Vols. 1–6, 2361–2387, Springer, New York, 2012.

[Young] https://en.wikipedia.org/wiki/Young%27s_interference_experiment

[Za13] Zawadzki, P.: *Closed-form formula on quantum factorization effectiveness*, Quantum Inf. Process. **12** (2013), 97–108 Wikipedia.